



**HAL**  
open science

# Cartographie fine de la recombinaison, analyse des séquences locales et étude du déséquilibre de liaison chez le blé tendre (*Triticum aestivum*)

Benoît Darrier

► **To cite this version:**

Benoît Darrier. Cartographie fine de la recombinaison, analyse des séquences locales et étude du déséquilibre de liaison chez le blé tendre (*Triticum aestivum*). Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2016. Français. NNT : 2016CLF22764 . tel-01548835

**HAL Id: tel-01548835**

**<https://theses.hal.science/tel-01548835>**

Submitted on 28 Jun 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Ecole Doctorale Sciences de la Vie, Santé, Agronomie, Environnement*  
N° d'ordre : 706

Présentée à l'université Blaise Pascal pour l'obtention du grade de  
**Docteur D'université** (Spécialité : Génétique)

Benoît DARRIER

---

Cartographie fine de la recombinaison, analyse des séquences locales et étude du déséquilibre  
de liaison chez le blé tendre (*Triticum aestivum*)

---

Soutenue le : 08/12/2016

Rapporteurs : V. Borde (Directrice de recherche Institut Curie)  
L. Duret (Directeur de recherche CNRS)  
Examineur<sup>o</sup>: E. Jenczewski (Directeur de recherche INRA IJPB)  
C. Moreno (Directrice de recherche INRA GenPhySE)  
C. Tatout (Professeur Université Blaise Pascal)  
F. Balfourier (Co-Directeur de thèse)  
P. Sourdille (Directeur de thèse)

Institut National de la Recherche Agronomique Site de Crouël UMR 1095 Génétique,  
Diversité et Ecophysiologie des Céréales



“The worst thing that can happen during the 1980’s is not energy depletion, economic collapse, limited nuclear war, or conquest by a totalitarian government. As terrible as these catastrophes would be for us, they can be repaired within a few generations. The one process in the 1980’s that will take millions of years to correct is the loss of genetic and species diversity by the destruction of natural habitats. This is the folly that our descendants are least likely to forgive us.”

From E.O Wilson (1985, from *The biological diversity crisis: a challenge to science*. *Issues Science Technology* 2: 20-25)



## Résumé

Mieux connaître les facteurs qui gouvernent l'apparition des événements de recombinaison (crossing-overs ; CO) chez le blé tendre (*Triticum aestivum* L.) est primordial car ce processus est l'outil principal du sélectionneur pour permettre le brassage de la diversité génétique et l'introgession de régions d'intérêt dans des variétés agronomiques. L'utilisation de techniques de cytogénétique développées sur l'orge a permis de comparer la mise en place de la synapse lors de la méiose chez des lignées de blé tendre délétées de tout ou partie du bras long du chromosome 3B et qui avaient été préalablement montrées comme présentant un nombre de chiasmas réduit par rapport à la variété euploïde. Les analyses cytogénétiques couplées à des études bioinformatiques de la séquence montrent que le synapsis a lieu quasiment normalement chez les mutants et que la délétion de certains gènes connus comme impactant le déroulement de la méiose pourrait expliquer le phénotype observé. De plus, le développement de ressources génomiques (SNP, séquence) à destination des sélectionneurs a permis la réalisation de cartes génétiques haute densité des 21 chromosomes ancrées sur la séquence du génome. Tous les chromosomes montrent le même profil de recombinaison avec un accroissement dans les parties distales et une réduction drastique dans les parties centromériques et péri-centromériques. L'exploitation de plus de 250 CO localisés dans des fenêtres de moins de 25 kb sur le chromosome 3B utilisé comme modèle pour l'étude de l'impact de la séquence sur la recombinaison, montre que les profils de recombinaison ancestrale et actuelle sont conservés et que les CO ont lieu préférentiellement dans les parties promotrices des gènes exprimés en méiose ce qui suggère que la conformation chromatinienne influence la recombinaison. Finalement, ces données ont aussi été l'opportunité de détecter des motifs liés à la recombinaison qui présentent des similarités avec celui ciblé par la protéine PRDM9 qui conduit à la recombinaison chez l'humain. Cela suggère que les mécanismes de contrôle de la recombinaison sont conservés chez les eucaryotes.

Mots clés : recombinaison ; méiose ; *Triticum aestivum* ; blé tendre ; crossing-over ; cartes génétiques ; cytogénétique.



## Abstract

Better knowledge of the factors that drive recombination (crossovers; COs) in bread wheat (*Triticum aestivum* L.) is of main interest since this process is the main tool allowing breeders to admix the genetic diversity and to introgress regions of interest in agronomic varieties. We used cytogenetic techniques previously developed on barley to compare the establishment of synapsis during meiosis in deletion lines missing part or whole of the long arm of chromosome 3B of bread wheat and which were previously shown as having a reduced chiasmata number compared to euploid varieties. Cytogenetic analysis combined with bioinformatics studies showed that the synapsis occurs almost normally in mutants and that deletion of some genes known to impact meiosis behavior may explain the observed phenotype. In addition, development of genomic resources (SNPs, sequence) for wheat breeders allowed simultaneous elaboration of high density genetic maps for the 21 chromosomes anchored on genome sequence. All chromosomes present the same recombination pattern with an increase of recombination in the distal parts and reduction in centromeric/pericentromeric regions of the chromosomes. Analysis of more than 250 COs mapped in windows lower than 25kb located on chromosome 3B used as model to study the impact of sequence features on recombination showed that current and ancestral recombination patterns are conserved and that COs preferentially occur in the promoter part of gene expressed in meiosis suggesting that chromatin conformation impacts recombination. Finally, these data were the opportunity to detect recombination-associated motif which presents similarities with the motif targeted by the PRDM9 protein driving recombination in human. This suggests that the control of recombination mechanisms is conserved among eukaryotes.

Key words: recombination; meiosis; *Triticum aestivum*; bread wheat; crossovers; genetic maps; cytogenetics.





## Remerciements

Je remercie les membres du jury Valérie Borde, Laurent Duret, Christophe Tatout, Carole Moreno et Eric Jenczewski d'avoir accepté d'évaluer mes travaux de recherche. Je tiens aussi à remercier particulièrement les membres de mon comité de thèse Christine Mézard, Bertrand Servin, Anne-Marie Chèvre, Frédéric Choulet, et Vincent Ranwez qui ont permis par leurs commentaires et leur implication la bonne conduite et la bonne réalisation de mes travaux.

Je souhaite remercier aussi mon encadrant direct : Pierre Sourdille. Merci à toi pour ces 3 années, certes très actives, mais riches d'expériences et de souvenirs (même si on se serait bien passé du problème à Los Angeles !). Merci de m'avoir laissé conduire mes travaux avec liberté et de m'avoir guidé dans mes réflexions. Merci aussi de ton implication dans la réalisation de cette thèse et du temps consacré. Je tiens aussi à te remercier de ton soutien et de ton éclaircissement dans mes problèmes éthiques et moraux qui m'ont permis de faire face. Merci pour tout !

Je souhaite remercier aussi mon co-encadrant François Balfourier pour son aide dans la réalisation de mes travaux sur le déséquilibre de liaison.

Je voudrais aussi remercier Isabelle Colas mon encadrante lors de mon court (et hyperactif !) passage en Ecosse. Merci pour ton initiation à la cytologie et merci pour la découverte de ta passion. Merci pour le temps que tu as pu m'accorder (bien que tout ne se soit pas passé comme prévu !). Merci pour ces souvenirs et ces belles images. J'ai découvert une superbe discipline et redécouvert le contact avec les plantes. Merci aussi, de m'avoir permis de participer à la « Barley retreat » et merci à Robbie Waugh pour la bière !

Et merci à Alain pour avoir appris la biologie moléculaire ! Merci aussi à Philippe pour son aide avec TriAnnot et son apport en bioanalyse.

Merci aussi à l'équipe GeCO dans sa globalité : Marie Claire, Véronique, Bouzid, Ludo, Robin, Marion, Fatiha, Isabelle... pour tous les bons moments.

Merci aux personnes de la plateforme, Charles, Anthony, Amélie, Véro, Lydia, Gégé et Delphine.

Je veux aussi remercier Frédéric Choulet pour ses conseils et sa vision des choses qui ont pu m'apporter des éclaircissements et m'aider en bioinformatique.

Merci à Josquin, Lise et Aurélie pour leurs conseils et leur aide quand je suis arrivé et merci de m'avoir permis de mettre le pied à l'étrier aussi rapidement et surtout merci pour les after work !

Un grand Merci aussi à Ambre-Aurore, Hélène et Jo ! Un grand merci à vous pour toutes ces parties de Ghost ! Merci aussi pour les moments de détente et les barres de rires. Merci pour tout ! Je retiendrai quand même que : Moi je peux me gratter le dos !

Et bien évidemment 3 ans c'est l'occasion de faire énormément de rencontre et il est impossible de citer ou de se souvenir de tout le monde. Alors je vais finir par dire merci à toutes les personnes qui ont jalonné ces trois années, merci à vous, merci à tous les Pierre, merci à Karine, merci à Nico, merci à .... Tous !



<b>INTRODUCTION</b> .....	<b>1</b>
<b>I. LA MEIOSE</b> .....	<b>4</b>
1. LA PROPHASE I :.....	5
A. <i>Leptotène / Leptonema (Figure 6)</i> .....	5
B. <i>Zygotène / Zygonema (Figure 7)</i> .....	6
C. <i>Pachytène / Pachynema (Figure 8)</i> .....	6
D. <i>Diplotène / Diplonema (Figure 9)</i> .....	7
E. <i>Diacinèse Figure 10</i> .....	7
F. <i>Fin de la méiose</i> .....	7
2. LE COMPLEXE SYNAPTONEMAL.....	8
3. L'APPARIEMENT DES CHROMOSOMES.....	10
4. SPO11 ET LA FORMATION DES CASSURES DOUBLE-BRIN.....	11
5. LOCALISATION DES CASSURES.....	13
6. RESOLUTIONS DES CASSURES.....	13
<b>II. LA RECOMBINAISON ET LE DESEQUILIBRE DE LIAISON</b> .....	<b>15</b>
1. METHODES DE DETECTION.....	16
2. REPARTITION ET FACTEURS DE VARIATION DES CO.....	18
3. POINT CHAUD DE RECOMBINAISON.....	19
4. CARACTERISTIQUES DES POINTS CHAUDS.....	19
5. EVOLUTION DES POINTS CHAUDS.....	21
6. DESEQUILIBRE DE LIAISON.....	22
<b>III. LE BLE TENDRE</b> .....	<b>25</b>
1. GENERALITES.....	25
2. ANALYSES PRELIMINAIRES DE LA RECOMBINAISON CHEZ LE BLE.....	26
3. ANALYSE DE LA RECOMBINAISON SUR LE CHROMOSOME 3B.....	26
4. UN PARTITIONNEMENT QUI N'EST PAS LIMITE A LA RECOMBINAISON.....	27
5. CONTROLE DE LA RECOMBINAISON HOMEOLOGUE.....	28
6. ETUDE DES POINTS CHAUDS DE RECOMBINAISON.....	30
7. ANALYSE DU DESEQUILIBRE DE LIAISON CHEZ LE BLE.....	30
<b>I. OBJECTIFS DE LA THESE</b> .....	<b>31</b>
<b>II. PARTIE I – CYTOGENETIQUE :</b> .....	<b>34</b>
1. TRANSFERT DES PROTOCOLES DE CYTOGENETIQUE DE L'ORGE VERS LE BLE.....	34
2. RECHERCHE DE GENES IMPLIQUES DANS LE COMPORTEMENT MEIOTIQUE SUR LE CHROMOSOME 3B.....	36
<b>LOCATION AND IDENTIFICATION OF GENES IMPACTING CHIASMATA NUMBER ON CHROMOSOME 3B OF BREAD WHEAT</b> .....	<b>37</b>
2. PRINCIPALES CONCLUSIONS.....	59
3. PERSPECTIVES.....	60

<b>III. PARTIE II – ANALYSE DE LA RECOMBINAISON AU NIVEAU DU GENOME ENTIER .....</b>	<b>61</b>
1. INTRODUCTION .....	61
<b>HIGH THROUGHPUT SNP DISCOVERY AND GENOTYPING IN HEXAPLOID WHEAT .....</b>	<b>62</b>
2. PRINCIPALES CONCLUSIONS .....	89
3. RESULTATS COMPLEMENTAIRES .....	90
A. <i>La cartographie du projet BreedWheat</i> .....	90
B. <i>Chinese SpringX Renan : PolyHigh</i> .....	91
C. <i>Biais de ségrégation</i> .....	92
D. <i>Analyse de la recombinaison génome entier</i> .....	93
4. DISCUSSION SUR LES RESULTATS COMPLEMENTAIRES .....	93
5. CONCLUSION .....	96
<b>IV. PARTIE III : ETUDE DES PARAMETRES QUI IMPACTENT LA DISTRIBUTION DE LA RECOMBINAISON SUR LE CHROMOSOME 3B ET L'ENSEMBLE DU GENOME .....</b>	<b>97</b>
1. INTRODUCTION .....	97
<b>HIGH-RESOLUTION MAPPING OF CROSSOVER EVENTS IN THE HEXAPLOID WHEAT GENOME. ....</b>	<b>98</b>
2. PRINCIPALES CONCLUSIONS .....	128
<b>V. CONCLUSION GENERALE ET PERSPECTIVES .....</b>	<b>129</b>
<b>VI. BIBLIOGRAPHIE.....</b>	
<b>ANNEXE.....</b>	
A. <i>Cartographie génome entier du blé tendre</i> .....	
B. <i>Script de traitement des données de génotypage</i> .....	
C. <i>Cartographique génétique.</i> .....	
D. <i>Population Chinese Spring X Renan</i> .....	
E. <i>Population ITMI</i> .....	
F. <i>Projection des cartes et addition de marqueurs par bioinformatique</i> .....	







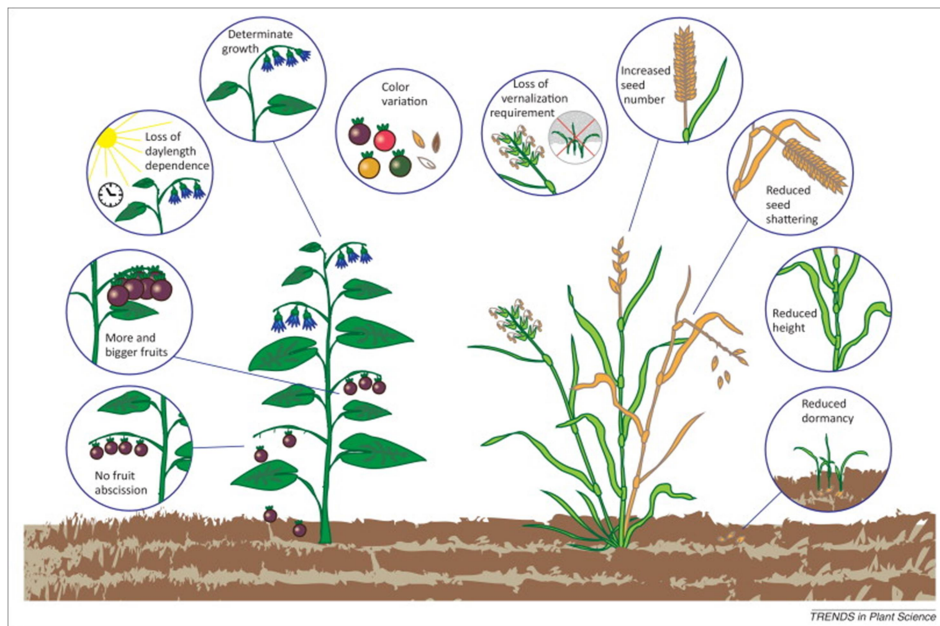


Figure 1 : **Caractères de domestication des plantes cultivées actuelles (d'après [1])**

Représentation schématique des grandes lignées de plantes angiospermes (eudicotylédones, à gauche, monocotylédones, à droite) à partir desquelles les cultures actuelles sont originaires, avec les caractères typiques des espèces sauvages représentés. Les caractères convergeant évolués présents dans diverses cultures domestiquées sont représentés dans les cercles.

La convergence des caractères phénotypiques des céréales et due aux besoins humains similaires pour toutes les cultures. On pourra noter, une volonté de maximiser le rendement qui va créer une pression de sélection sur le nombre et la taille des parties de la plantes utilisées et qui va conduire à la suppression du phénotype sauvage. De plus, le mécanisme de dispersion du fruit sera éliminé pour réduire les pertes de rendement. Ensuite, le changement des zones de culture nécessite souvent un changement simultané dans la dépendance à la photopériode et la vernalisation. La réduction de la dormance des graines est aussi nécessaire pour la synchronisation de la germination. Finalement, les préférences esthétiques entraînent souvent des adaptations convergentes. Un exemple très marquant est la couleur.

## Introduction

---

Au cours du temps, les sociétés humaines ont réalisées une transition majeure il y a environ 12 000 ans grâce à la domestication des plantes et des animaux [2,3] qui a permis le passage d'un mode de vie nomade vers une sédentarisation avec la naissance de la pratique de l'agriculture et de l'élevage [2]. Si la domestication a profondément changé le mode de vie de l'homme, elle a aussi eu un impact important sur les organismes domestiqués. Ainsi, chez les plantes, des modifications morphologiques essentielles comme la taille et le nombre de graines ou encore leur maintien sur épis ont été décrits [2–4] (Figure 1).

Il est proposé par certains auteurs que la domestication par l'homme de certaines espèces aurait favorisé la polyplöidie [5] car celles-ci ne sont connues qu'à l'état cultivé. C'est le cas par exemple du blé tendre et du colza [6]. La présence de plusieurs génomes est due soit à la multiplication de l'information déjà présente (auto-polyplöide), soit à l'addition d'informations par hybridation provenant d'espèces apparentées (allo-polyplöide) [7]. Leur génome est constitué d'une juxtaposition de génomes plus simples. C'est par exemple le cas du fraisier ( $2n=8x$ ), de la pomme de terre ( $2n=4x$ ), du blé tendre ( $2n=6x$ ), du colza ( $2n=4x$ ). Aujourd'hui, la majeure partie des plantes cultivées est polyplöide [8].

La sélection génétique des espèces cultivées ainsi que leur domestication conduisent souvent à un goulot d'étranglement génétique à cause de la contribution d'un nombre limité d'individus à la naissance et à l'amélioration de l'espèce concernée [3,4,9,10]. Ainsi, la standardisation des variétés qui est nécessaire à l'agriculture intensive, contribue dans le même temps à la réduction de la diversité [11–13]. Malgré les avantages d'un système standardisé et intensif, la faible base génétique des plantes cultivées va participer au développement de maladies à large échelle (Cf Mildiou en Irlande 1840) [14].

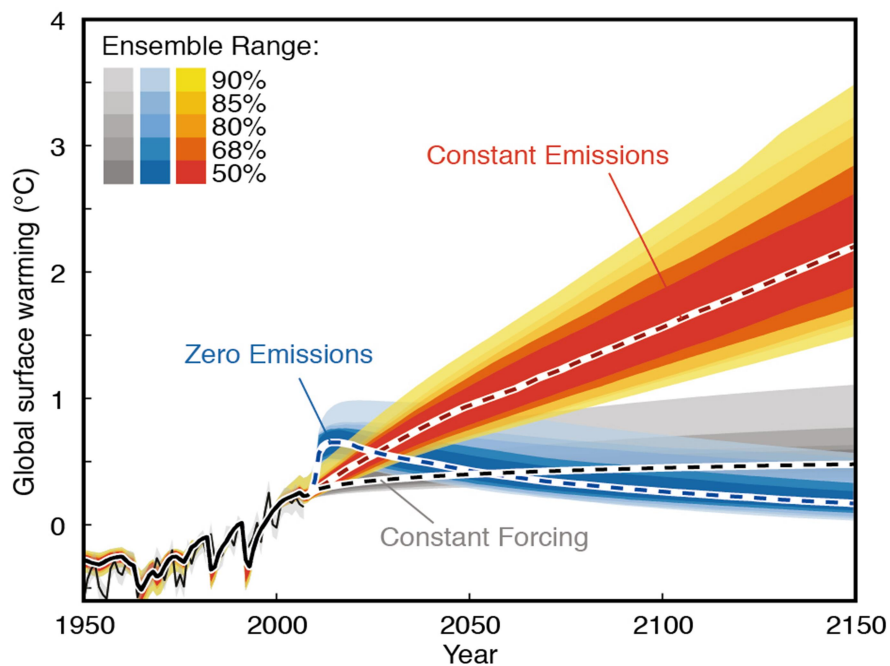


Figure 2 : Réchauffement climatique, quelles prévisions ? (d'après [15])

Projection de l'augmentation globale de la température basée sur le modèle du bilan énergétique du cycle du carbone pour l'évaluation des gaz à effet de serre inducteurs du changement climatique (MAGICC- Model for the Assessment of Greenhouse Gas-Induced Climate Change) pour une composition atmosphérique constante ('constant forcing', gris), émissions constantes (rouge) et zéro émissions futures (bleu) à partir de 2010, avec l'estimation de l'incertitude. La ligne noire fine et l'ombrage indiquent le réchauffement et l'incertitude observée.

Les systèmes intensifs induisent également l'épuisement, la pollution et la salinisation des sols et des réserves d'eau douce. De plus, les activités anthropiques contribuent au processus de réchauffement climatique global (Figure 2). Cependant, ces systèmes ont prouvé leur efficacité avec une augmentation régulière des rendements depuis le XIX<sup>ème</sup> siècle couvrant ainsi les besoins alimentaires des populations.

Aujourd'hui, la FAO (Organisation des nations unies pour l'alimentation et l'agriculture) estime que la démographie humaine approchera les 9 milliards d'individus en 2050 et que pour couvrir la demande alimentaire concomitante, les rendements devront augmenter de 60% par rapport aux rendements de 2005-2007 [13]. Si jusqu'à présent, la réponse à une demande croissante s'est faite par une augmentation des surfaces de production, cette option n'est plus possible du fait d'une réduction des terres arables due à une urbanisation galopante. De plus, vient se greffer la nécessité de réduire les intrants (engrais, pesticides, fongicides, eau) pour diminuer les effets secondaires dus à une production intensive. Pour répondre aux problématiques, il est nécessaire d'accélérer l'adaptation de nos plantes cultivées vers un système de production écologiquement intensive et durable pour assurer la pérennité des rendements et garantir la sécurité alimentaire. Différentes stratégies pour l'accélération de la création variétale peuvent être envisagées, mais sont dépendantes (1) de la diversité génétique disponible ( Figure 3) et (2) de la capacité à insérer uniquement les caractères d'intérêt dans un nouveau génotype. Ce transfert ciblé d'allèles d'intérêt n'est pas toujours aisé.

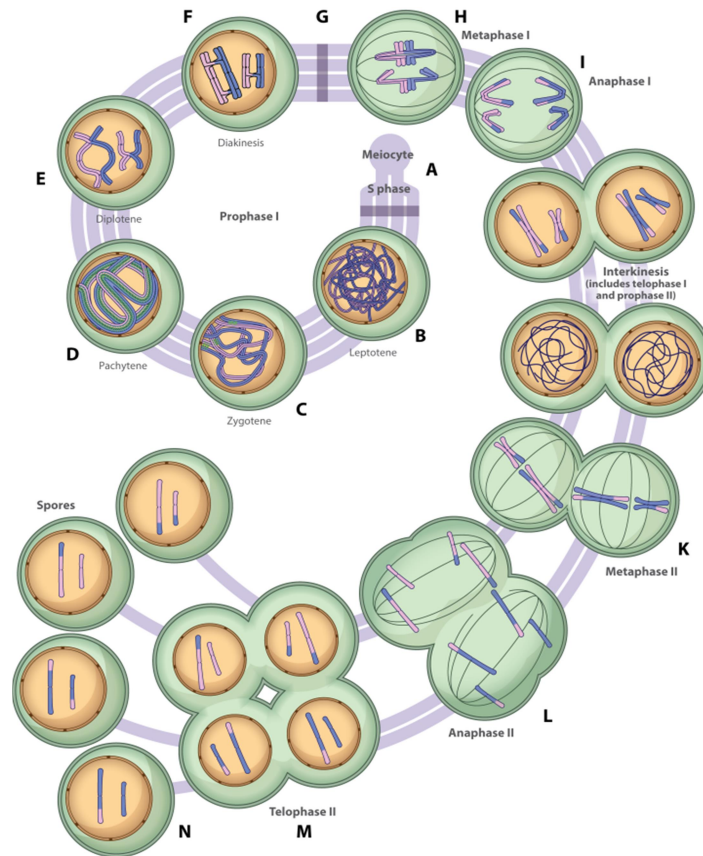
Aujourd'hui, les techniques modernes de génie génétique permettent d'introduire de manière rapide et héritable des gènes d'intérêt dans le génome d'une variété ciblée. Cependant, la manipulation de l'information génétique reste difficilement acceptable par certaines sociétés. De même, elle implique de pouvoir contrôler la reproduction de la plante afin de proscrire la dissémination du/des caractère(s) introduit(s) dans l'écosystème.



Figure 3 : **Diversité du blé tendre**

Représentation de la diversité du blé tendre au niveau de l'épi. (Copyright : INRA)

Une autre solution serait d'accentuer l'effort de sélection en exploitant des croisements inter et/ou intra-spécifiques et d'utiliser les moyens naturels de brassage génétique comme la recombinaison pour l'optimisation des schémas de sélection et la création de nouvelles variétés innovantes. Mes travaux de thèse s'inscrivent dans ce contexte particulier et ont pour objectif de contribuer à la compréhension de la localisation des crossing-overs et la répartition du déséquilibre de liaison chez une espèce de grande culture d'intérêt socio-économique mondial : le blé tendre.



Mercier R, et al. 2015. Annu. Rev. Plant Biol. 66:297–327

Figure 4 : Déroulement Méiose (d'après [16])

La méiose. (A) La préméiose inclut la différenciation cellulaire et la phase S méiotique. (B) Leptotène, les axes des chromosomes se forment et la recombinaison débute. (C) Zygotène, les chromosomes s'apparient (synapsis) à l'aide de la polymérisation du complexe synaptonémal et la recombinaison progresse. (D) Pachytène, le synapsis est terminé et la recombinaison continue de se dérouler. (E) Diplotène, le complexe synaptonémal se dissocie et les chromosomes homologues sont connectés par les crossing-overs. (F) Diakinesis, la condensation des chromosomes a lieu et les bivalents peuvent être visibles. (G) La Prophase I se termine et l'enveloppe nucléaire disparaît. (H) Métaphase I, le cytosquelette apparaît et les bivalents s'alignent sur la plaque métaphasique. (I) Anaphase I, le relâchement de la cohésion entre chromatides sœurs permet la migration des chromosomes aux deux pôles. La cohésion péricentromérique est spécifiquement protégée. (J) Intercinèse, deux noyaux se forment et les chromosomes se décondensent très brièvement. Cette étape englobe la télaphase I et la prophase II. Chez les monocotylédones, la division du cytoplasme ce fait avant que la méiose II ne commence ; chez les dicotylédones, la séparation ce fait seulement à la télaphase II. (K) Métaphase II, les deux cytosquelettes alignent les chromosomes sur deux plaques métaphasiques. (L) Anaphase II, les chromatides sœurs se séparent grâce au relâchement de la cohésion centromérique. (M) Télaphase II, les quatre noyaux se forment. (N) Cytocinèse, quatre spores haploïdes sont formés.

## I. La méiose

---

Malgré le coût énergétique plus élevé d'une reproduction sexuée dû à la fabrication des gamètes, la recherche de partenaires, etc. celle-ci s'est maintenue grâce à ses nombreux avantages tels que l'adaptation à différents types d'environnements permise par la variabilité génétique des individus [17].

Cette variabilité est en partie induite par la méiose, un type de division commun à tous les eucaryotes qui permet la formation des gamètes (Figure 4). Elle se compose de deux divisions successives qui assurent le passage d'une cellule diploïde, à quatre cellules haploïdes. Le niveau de ploïdie des individus sera ensuite rétabli par la fécondation. C'est lors de ces deux divisions successives que le brassage génétique se fait par le biais des crossing-overs et la répartition aléatoire des chromosomes [18].

Chez les plantes, la méiose ne présente aucun point de contrôle qui, en cas de problème de ségrégation des chromosomes, arrête la division comme chez la levure, ou déclenche l'apoptose chez les animaux [19]. Les différentes étapes de la méiose se discriminent aisément par observation cytologique grâce à des caractéristiques distinctes des chromosomes (Figure 5).



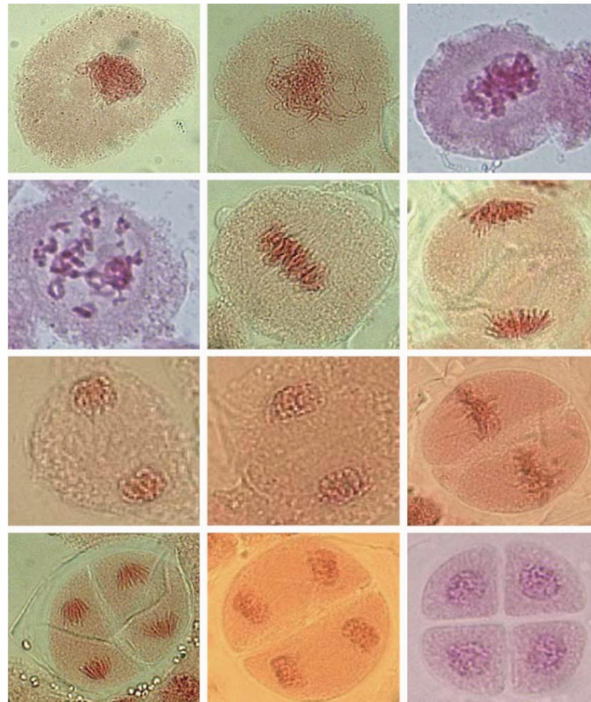


Figure 5 : La méiose du blé tendre (*Triticum aestivum*) (d'après[20])

(A-E) Prophase I avec (A) zygotène, (B) pachytène, (C) diplotène et (D) diacinèse. (E-G) représentent les stades restant de la première division de méiose avec (E) illustrant la métaphase I, (F) l'anaphase I et (G) la téléphase I. La seconde division de méiose est montrée de (H-L) avec (H) qui représente la prophase II, (I) métaphase II, (J) anaphase II, (K) téléphase II et (L) la formation de la tétrade.

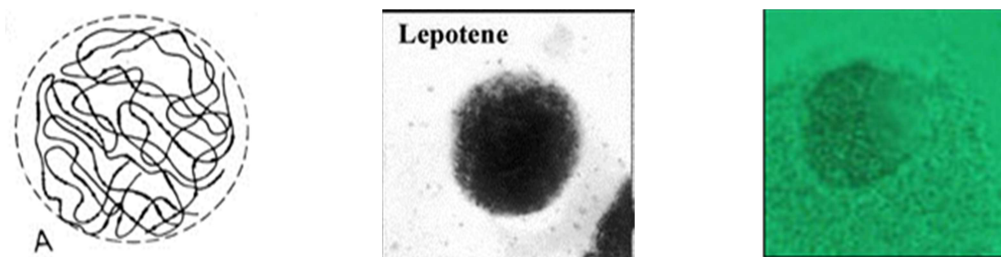


Figure 6 : Leptotène

De gauche à droite : Représentation schématique de l'étape de méiose (Eric Walravens, Athénée Fernand Blum de Schaerbeek, Bruxelles, Belgique). Image cytologique (*Lilium grandiflorum*) (copyright site de Radboud University Nijmegen). Image cytologique blé tendre : I. Colas.

## 1. La prophase I :

La prophase I suit directement la phase de réplication de l'ADN (Phase S) et est l'étape la plus longue de la méiose car elle est le siège de profonds changements de l'état chromatinien, ce qui a conduit à sa division en 5 phases distinctes :

### A. Leptotène / Leptonema (Figure 6)

Lors de cette première phase les chromosomes commencent à se condenser et deviennent visibles sous la forme de fins filaments. A ce stade, les observations réalisées au microscope électronique montrent que les chromosomes commencent à s'organiser en boucle le long de segments d'éléments axiaux [21]. C'est lors du leptotène que les cassures double-brin programmées de l'ADN, futurs points de départ de la recombinaison, sont générées. A la fin du leptotène, des structures particulières sont observées, appelées nodules précoces et nodules tardifs selon leurs tailles et moments d'apparition. Les premiers, généralement petits, distribués de manière irrégulière, correspondent probablement aux cassures double-brin de l'ADN et/ou à des débuts d'interactions entre chromosomes homologues. Les seconds, plus gros, présents en pachytène correspondraient à la future recombinaison [21,22]. Dans les deux cas les nodules contiennent des protéines impliquées dans la recombinaison [23].

Ensuite lors de la transition vers la phase zygotène les télomères s'attachent à l'enveloppe nucléaire par la plaque d'attache et forment une structure particulière appelée le bouquet télomérique [23,24]. Cette structure quasi-universelle et non statique est supposée favoriser l'appariement des chromosomes et éviter l'enchevêtrement des chromosomes (interlocks) [23,25].



Figure 7 : **Zygotène**

De gauche à droite : Représentation schématique de l'étape de méiose (Eric Walravens, Athénée Fernand Blum de Schaerbeek, Bruxelles, Belgique). Image cytologique (*Lilium grandiflorum*) (copyright site de Radboud University Nijmegen). Image cytologique blé tendre : I. Colas.

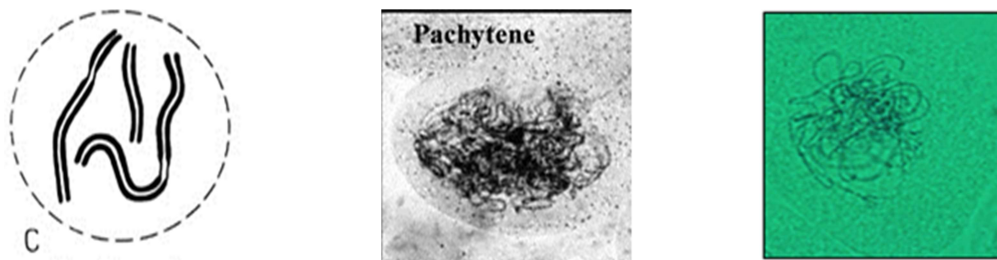


Figure 8 : **Pachytène**

De gauche à droite : Représentation schématique de l'étape de méiose (Eric Walravens, Athénée Fernand Blum de Schaerbeek, Bruxelles, Belgique). Image cytologique (*Lilium grandiflorum*) (copyright site de Radboud University Nijmegen). Image cytologique blé tendre : I. Colas.

## B. Zygotène / Zygonema (Figure 7)

Les chromosomes s'apparient (ou synapsis) aidés par la mise en place du complexe synaptonémal (ou synapton). Ce dernier est constitué d'éléments axiaux nommés éléments latéraux, et d'éléments centraux qui jouent le rôle d'une fermeture éclair pour rapprocher les deux chromosomes homologues. L'initiation de sa formation est espèce-dépendante mais se fait au niveau de plusieurs sites non aléatoires le long des chromosomes et tend à débiter préférentiellement par les régions télomériques favorisées par le bouquet. Cependant, chez les plantes, plusieurs sites d'initiation additionnels du synapsis sont présents le long des chromosomes et la position est probablement en lien avec la taille des chromosomes [22,25,26]. Les observations 3D chez la levure (*in vivo* et *in vitro* par EPI fluorescence) ont mis en évidence des mouvements très lents de petits segments de chromosomes durant le zygotène [27].

## C. Pachytène / Pachynema (Figure 8)

A ce stade, les chromosomes sont appariés alors que la polymérisation du complexe synaptonémal le long des chromosomes homologues s'achève. Cette phase est considérée comme la période active de l'appariement des chromosomes (pairing). Les chromosomes sont sous forme de filaments et la recombinaison génétique est achevée. Plusieurs évidences montrent qu'à ce stade certains chromosomes sont encore attachés au nucléole, mais aussi que des mouvements chromosomiques peuvent encore être observés. Cependant, les segments mobiles sont beaucoup plus longs et présentent un balayage lent de grandes étendues du noyau.



Figure 9 : **Diplotène**

De gauche à droite : Représentation schématique de l'étape de méiose (Eric Walravens, Athénée Fernand Blum de Schaerbeek, Bruxelles, Belgique). Image cytologique (*Lilium grandiflorum*) (copyright site de Radboud University Nijmegen). Image cytologique blé tendre : I. Colas.



Figure 10 : **Diakinèse**

De gauche à droite : Représentation schématique de l'étape de méiose (Eric Walravens, Athénée Fernand Blum de Schaerbeek, Bruxelles, Belgique). Image cytologique (*Lilium grandiflorum*) (copyright site de Radboud University Nijmegen). Image cytologique blé tendre : I. Colas.

#### D. Diplotène / Diplonema (Figure 9)

Pendant cette phase, les chromosomes se décondensent et prennent un aspect plus diffus. Les homologues se séparent de manière longitudinale (désynapsis). De plus, le complexe synaptonémal se démantèle et des faisceaux de microtubules s'accrochent aux niveaux des kinétochores des chromosomes. Par contre, les chromosomes restent attachés en des points de contact particuliers appelés les *chiasmata* (singulier : *chiasma*) qui sont la structure cytologique qui caractérise les crossing-overs et dont le nombre est variable et espèce-dépendante.

#### E. Diacinèse Figure 10

Enfin les chromosomes se re-condensent et chaque bivalent peut-être observé distinctement. Le complexe synaptonémal est complètement démantelé et les *chiasmata*, toujours visibles, assurent le lien physique entre chromosomes.

#### F. Fin de la méiose

Lors de la **Métaphase I**, les bivalents s'alignent le long de la plaque équatoriale de manière à ce que les chromosomes de chaque paire fassent face à chaque pôle de la cellule, de façon aléatoire. Ensuite, en **Anaphase I**, il y a séparation et migration vers chaque pôle des chromosomes avec un maintien de la cohésion des centromères. La répartition aléatoire des éléments d'une paire va permettre un second brassage génétique (brassage inter-chromosomique). Vient ensuite la **Télophase I** lors de laquelle il y a formation de 2 cellules haploïdes qui vont posséder un set aléatoire de chromosomes à deux chromatides. A ce stade, les enveloppes nucléaires peuvent parfois réapparaître et les chromosomes se décompacter légèrement selon les espèces. Par la suite une seconde division sur le même modèle que la première se produit. Cependant, du fait de la compaction des chromosomes déjà établie, la prophase est beaucoup plus rapide.

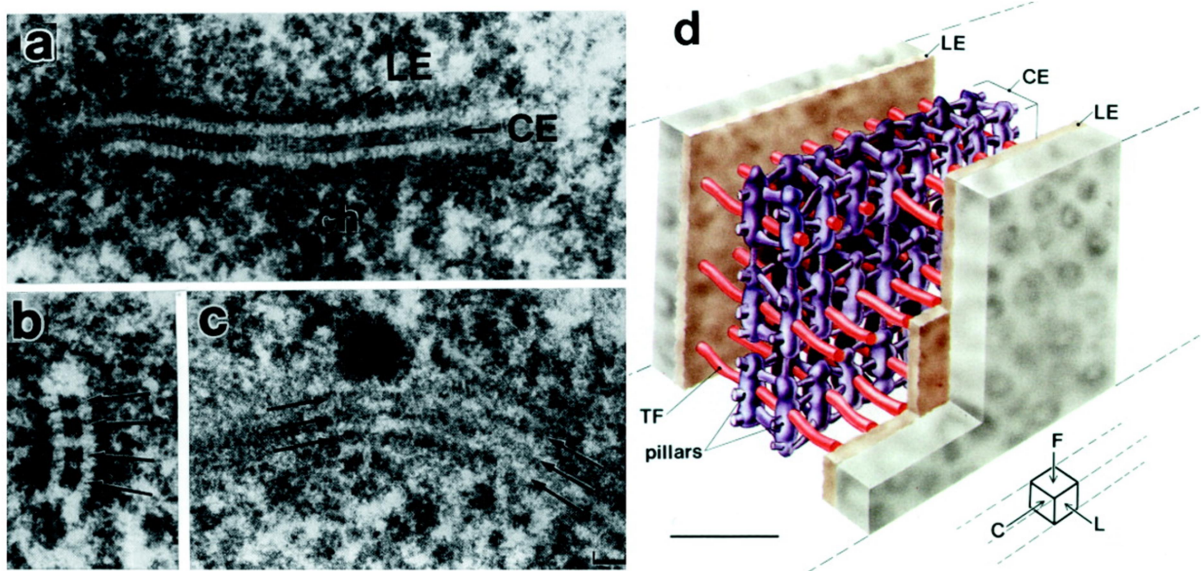


Figure 11 : Le Complexe Synaptonemal (d'après [22])

Le complexe synaptonémal (a) coupe longitudinale du complexe synaptonémal de *Blaps cribrosa* avec un élément central (EC) en forme d'échelle et deux éléments latéraux (EL) entourés de chromatine (ch). (b) coupe transversale qui montre quatre couches de l'EC. (c) coupe latérale avec trois couches d'EC indiquées par les flèches (la boule foncée sur le dessus des couches est un nodule de recombinaison). (d) modèle en trois dimensions de la région centrale chez *Blaps cribrosa*. L'EC multicouche est en violet, les deux EL flanquant sont en jaune pâle (modélisé en plaques), les filaments transversaux (FT) traversant la région centrale en rouge, et la chromatine environnante est en gris. Dans la région centrale, une unité structurelle provisoire peut être identifiée consistant en un FT avec deux piliers correspondant. Des ponts fibreux rejoignent les deux piliers liés au même FT, ainsi que sur les piliers FT adjacents pour former une couche de type échelle EC. Les couches individuelles sont reliées par d'occasionnelles fibres verticales reliant les piliers les uns sur les autres et en maintenant approximativement les couches. Angle de vue : F, frontal, L, latéral, C, vue en coupe transversale. Echelle = 50 nm.

Les différents stades, **Métaphase II**, **Anaphase II**, **Télophase II** sont similaires à la première division de méiose. Cependant, lors de cette seconde division, il y a scission des centromères pour former 4 cellules haploïdes contenant des chromosomes à une 1 chromatide. A la suite de la **Télophase II** ces quatre cellules sont accrochées sous forme de **Tétrades** qui après différenciation, dans le cas des plantes, formeront les grains de **Pollen**.

## 2. Le complexe synaptonémal

Comme présenté précédemment le complexe synaptonémal est une structure spécifique de la méiose conservée au sein de l'évolution probablement pour son rôle structural. En effet, les protéines qui le composent possèdent une très faible homologie de séquence entre les différentes espèces, mais une organisation structurale commune est maintenue. Aussi, il est probable que la structure de ces protéines soit soumise à sélection par leur interaction avec d'autres protéines [25].

Ainsi le complexe synaptonémal se présente sous une forme d'échelle composée d'éléments axiaux constitués d'une structure protéique continue le long de chaque chromosome et d'éléments centraux (Figure 11). Différentes études biochimiques ont pu montrer qu'au stade pachytène les éléments axiaux sont constitués de protéines structurales mitotique ou méiose-spécifiques comme des topo Isomérase II, des condensines, des cohésines et des protéines associées aux cohésines.

Dans beaucoup d'organismes, une importante partie de l'axe est constituée d'une ou plusieurs protéines à domaines HORMA, impliquées dans l'association à la chromatine [28]. Les échelons pour leur part sont composés d'un dimère de protéines coiled-coil *Zip1* (*SCP1* /*SYCP1* /*C(3)G* /*SYP-1* /*ZEP1*) chez tous les organismes (Figure 12). Ils ne sont cependant pas les seuls constituants des éléments centraux mais sont associés avec d'autres protéines en fonction des organismes [25].



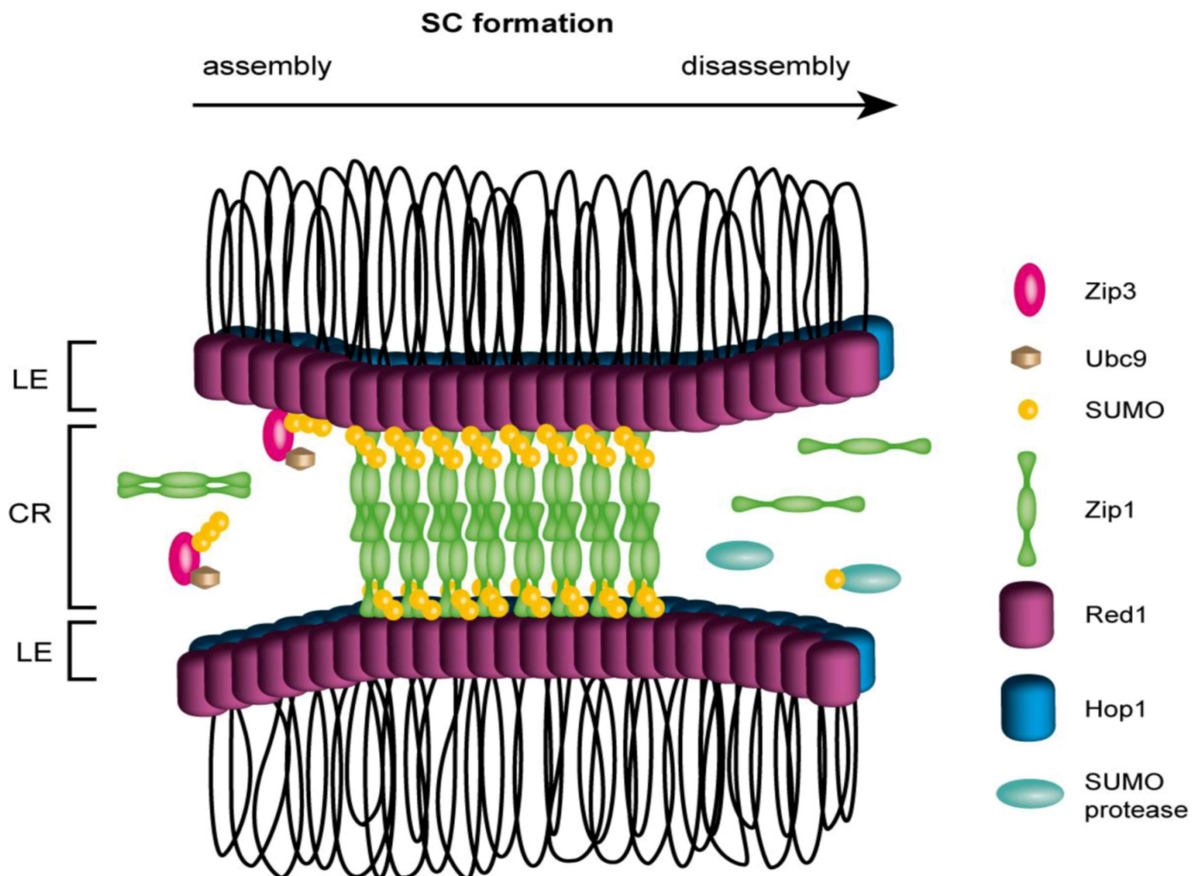


Figure 12 : La formation du complexe Synaptonémal (CS) (d'après [29])

Lors de l'assemblage du CS, *Zip3* recrute *Ubc9* et des protéines SUMO (small ubiquitin-like modifier) au sein du siège du synapsis, facilitant ainsi la formation des chaînes SUMO et de la conjugaison de SUMO à d'autres protéines (telles que *Red1*). Les dimères de *Zip1* se polymérisent le long des éléments latéraux (LE) où ils peuvent se lier à des chaînes *Red1* et les chaînes SUMO menant à la création de la région centrale (CR). Le démontage du CS pourrait s'effectuer par la dissociation des chaînes SUMO conjuguée à l'action de SUMO protéases ou tout autre mécanisme antagoniste. Pour plus de clarté, *Red1* et *HOP1* sont illustrés le long des LE, même si la répartition exacte des *HOP1*, *Red1* et *Red1*-sumoylé reste inconnue.

Du fait de la faible conservation des séquences, il est difficile de parvenir à trouver les gènes homologues de ceux identifiés chez les espèces modèles comme la levure, ou chez les espèces dont la séquence du génome est disponible (comme *Arabidopsis thaliana*) [30]. Néanmoins, par l'identification de mutants qui présentent une réduction de la fertilité, un profil asynaptique assorti d'une réduction des chiasmas, le locus *ASY1* a été isolé chez *Arabidopsis*. Ainsi, le clonage d'*ASY1* a montré des similarités dans son domaine HORMA en N-terminal avec un élément axial de la levure *HOP1*. Les études de cytogénétique et de microscopie électronique ont montré qu'*ASY1* est localisé dans les régions chromatiniennes à proximité des éléments axiaux et latéraux [31,32].

Malgré les avancées faites dans la détermination de cette structure particulière beaucoup d'éléments restent inconnus et la/les fonction(s) du complexe synaptonémal reste(nt) aujourd'hui mal comprise(s). Néanmoins, celui-ci semble jouer à la fois un rôle à l'échelle du chromosome entier et au niveau des sites de recombinaison.

En effet, la plupart des mutants de protéines impliquées dans les éléments axiaux ou centraux du complexe synaptonémal présentent un défaut du synapsis et une réduction voire une absence totale de recombinaison. Aussi, chez la levure ou chez *Arabidopsis* sa formation semble être conditionnée par la formation des cassures double-brin de l'ADN, générées par *SPO11*, qui vont initier la recombinaison [33].

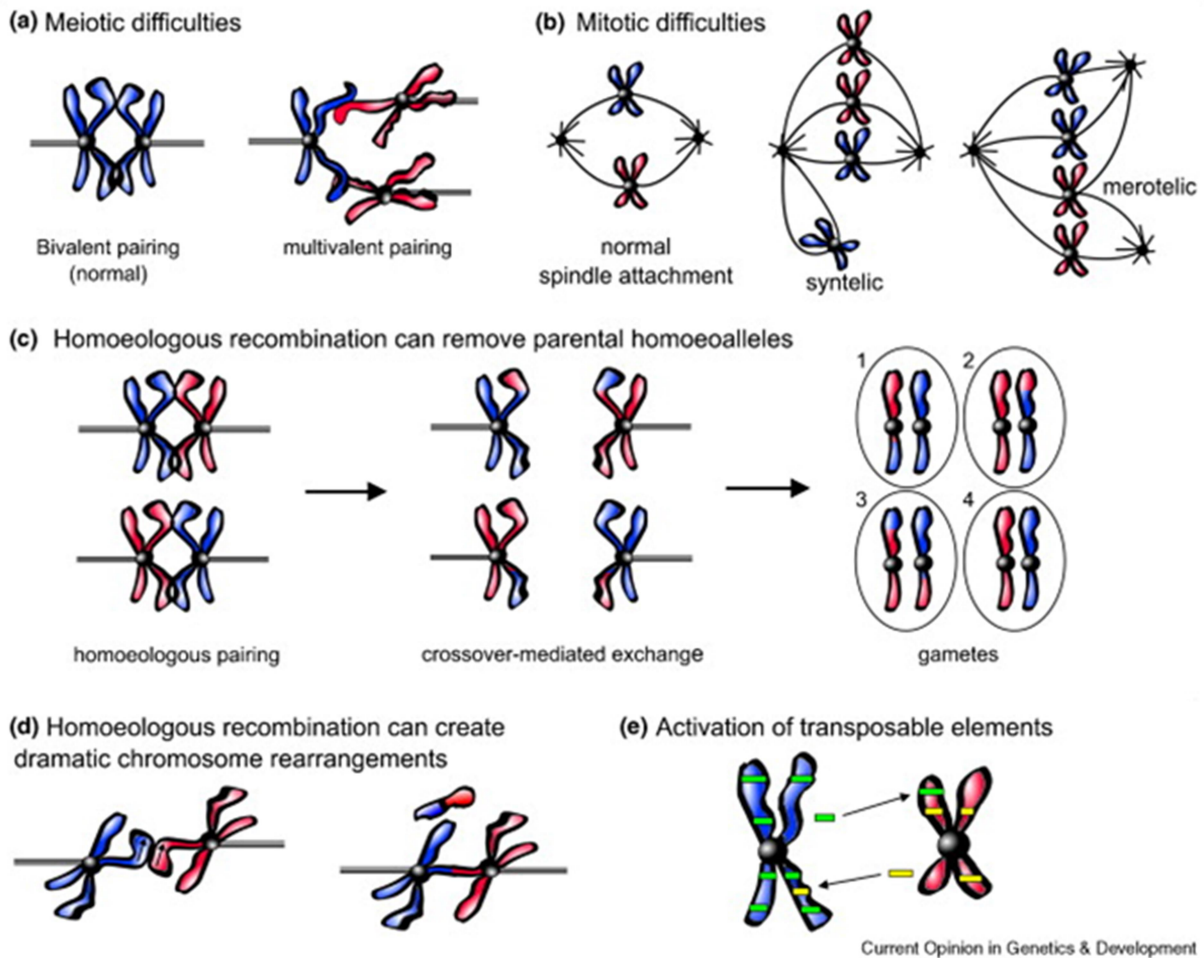


Figure 13 : Restructuration du génome par la polyploïdie (d'après [34])

La polyploïdie peut induire des changements dans la structure du génome par le biais d'une variété de mécanismes moléculaires. **(a)** et **(b)** les organismes polyploïdes peuvent rencontrer des problèmes durant la méiose et la mitose ce qui peut conduire à la formation des gamètes ou cellules aneuploïdes et à de possibles réarrangements du génome. **(c)**, les génomes polyploïdes possèdent beaucoup d'homologie de séquence entre chromosomes, ce qui peut engendrer des recombinaisons alléliques non homologues particulièrement entre chromosomes homéologues chez les allo polyploïdes. **(d)** La recombinaison homéologue peut aussi créer des chromosomes anormaux quand il préexiste des réarrangements entre les chromosomes homéologues. Dans cet exemple, le chromosome homéologue possède une inversion dans la région distale du bras. Lors de la recombinaison, l'échange va créer un chromosome di-centrique et un acentrique.

### 3. L'appariement des chromosomes

L'un des principaux défis lors de la méiose est la recherche du chromosome homologue afin de pouvoir former les crossing-overs et assurer une ségrégation correcte. Des observations par immuno-histologie chez les maïs [35] montrent que le co-alignement est presque complet avant le début de la formation du bouquet télomérique. Par contre, les forces qui dirigent ce co-alignement ne sont que partiellement comprises et restent à élucider. Cependant, il est suggéré que le mouvement des chromosomes pourrait permettre un contact initial par collision aléatoire.

Ces mouvements pourraient être aidés par la quasi-universelle augmentation de taille lors de la phase G1. Plusieurs modèles ont été proposés, incluant une recherche de contact ADN/ADN ou ADN/ARN ou la recherche de séquences d'appariement particulier.

Ces mouvements de chromosomes sont aussi requis pour que les cassures double-brin identifient la séquence homologue du chromosome partenaire dans le génome entier. En principe elle est trouvée par le mouvement des chromosomes et/ou grâce aux tentacules d'ADN libérés par la cassure double-brin qui peut rechercher à travers l'espace du noyau. Ces mouvements peuvent être créés par mouvement thermique, remodelage de la chromatine, métabolisme ADN/ARN, par la structure du chromosome en prophase, et/ou au cours de la création des cassures double-brin générées par la reconnaissance de l'homologue et le co-alignement lui-même.[25,36].

Lors des phénomènes de polyploïdisation les similarités entre chromosomes homologues et homéologues peuvent engendrer des problèmes d'appariement, de ségrégation, des gamètes aneuploïdes et des réarrangements chromosomiques (Figure 13). Ces phénomènes sont notamment la conséquence d'appariements multivalents [37]. Ainsi les phénomènes de polyploïdisation nécessitent une régulation de la méiose permettant la formation de gamètes équilibrés afin d'obtenir le rétablissement de la fertilité ainsi que la mise en place d'une barrière reproductive nécessaire à la spéciation.

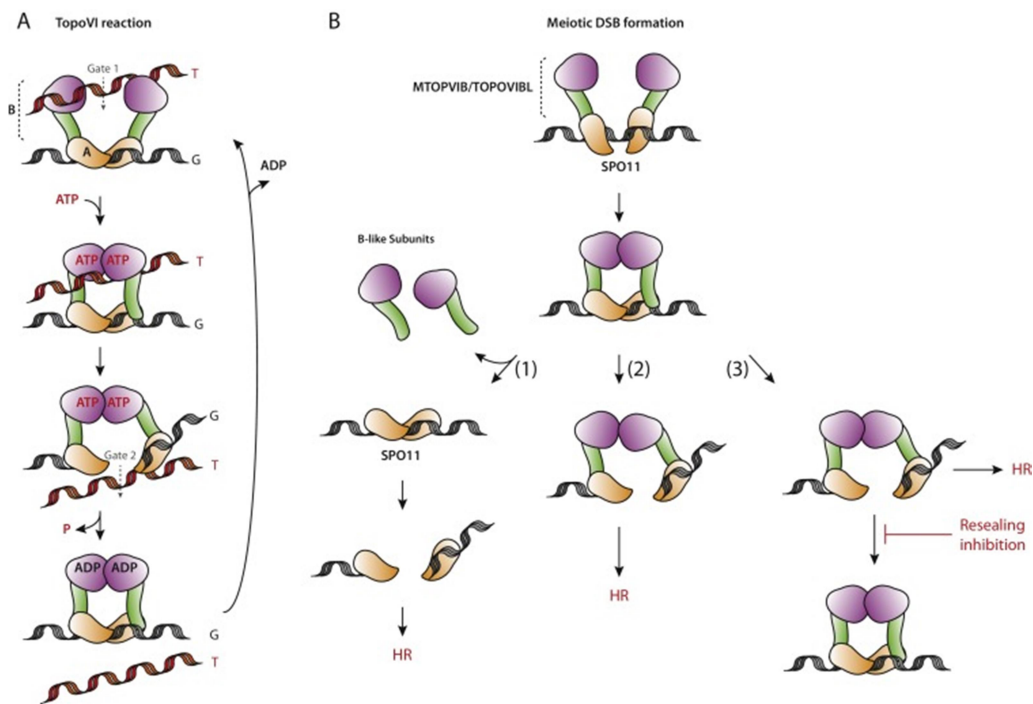


Figure 14 : **Modèle pour la formation des cassure double-brin(d'après [38])**

(A) Modèle dit «à deux portes» des TopoVI. Les TopoVI sont des enzymes hétérotétramériques ( $A_2B_2$ ) qui catalysent le passage d'un duplex d'ADN vers une cassure double-brin de l'ADN transitoire. La fixation de l'ATP et son hydrolyse par le domaine GHKL sont requis pour la dimérisation de la sous unité B, la capture du duplex d'ADN, la formation d'un site catalytique actif, l'ouverture et le clivage du duplex d'ADN et le passage du brin avec re-fermeture. Le segment présent dans la porte (G) est présenté en noir et le segment transféré (T) en rouge. La porte 1 où le segment T est capturé et la porte 2 dans laquelle le segment est transféré sont indiquées. (B) modèle proposé pour la formation des cassures double-brin méiotiques. La formation des cassures double-brin est catalysée par un dimère de *SPO11*, dont la formation est susceptible d'être médiée par des protéines méiotiques analogues à la *TopoVIB*. Une fois le dimère catalytique de *SPO11* formé, les sous-unités B peuvent être optionnelles (1). De manière alternative, la formation d'un complexe stable analogue à la *TopoVI* impliquant deux sous unités *SPO11* et deux sous-unités méiotiques *TopoVIB* pourrait être nécessaire pour couper l'ADN et déclencher la recombinaison homologue. La liaison de l'ATP (et l'hydrolyse) par les sous-unités analogues de la *TopoVIB* pourrait être nécessaire pour activer le dimère de *SPO11* et réaliser la formation des cassures double-brin de manière irréversible (2) ou réversible (3). Dans ce dernier cas, les activités supplémentaires seraient nécessaires pour empêcher la re-fermeture et pour permettre la recombinaison homologue.

Différentes études ont montré que des mécanismes de contrôle de l'appariement existent afin d'éviter les problèmes d'appariement de type multivalent, comme chez le blé hexaploïde où le gène *Ph1* (Pairing homoeologous 1) supprime les appariements entre les chromosomes homéologues. Un autre exemple est *PrBn* (Pairing Regulator in *B. napus*) chez le colza qui favoriserait les appariements homéologues

#### 4. SPO11 et la formation des cassures double-brin

Afin de pouvoir réaliser la recombinaison, des cassures double-brin de l'ADN doivent être créées au stade leptotène via la protéine *SPO11*. Cette dernière présente des similarités avec les sous-unités catalytiques des protéines Topo-isomérase VI qui sont des enzymes d'une importance capitale pour la régulation de la topologie de l'ADN durant différents processus clef comme la recombinaison. Elles vont notamment pouvoir couper, bouger, rejoindre l'ADN [38].

Le mécanisme de cassure double-brin est conservé à travers toutes les espèces et des homologues de *SPO11* ont été trouvés chez tous les eucaryotes (Figure 14). Chez *Arabidopsis*, deux homologues de *SPO11* sont décrits et requis pour la formation des cassures : *SPO11-1* et *SPO11-2* [38–41].

La suppression d'une de ces protéines ou une simple mutation du site catalytique suffit à empêcher la formation des cassures double-brin. Cela suggère qu'elles fonctionnent plutôt en hétéro-dimères. Ce mécanisme conservé et complexifié chez les plantes pourrait l'être d'avantage chez certaines plantes comme le riz chez qui d'autres *SPO11* additionnels ont été identifiés [33,42,43].

Tableau 1 : Ratio entre cassures double-brin et crossing-over (d'après [44])

Organism	Number of DSBs (# per homolog pair)	Number of CO (# per homolog pair)	Ratio DSBs/CO	% interfering CO	DSBs needed for pairing?
<i>S. cerevisiae</i>	160 (10)	90 (5.6)	1.8	75–85%	Y
<i>Mus musculus</i>	400 (20)	26 (1.3)	15	75–85%	Y
<i>Arabidopsis</i>	250–300 (50–60)	10 (2)	25–30	75–85%	Y
<i>Sordaria</i>	58 (8.3)	21 (3)	2.8	75–85%	Y
<i>Drosophila</i>	22 (5.5)	5 (1.2)	4.4	n.d.	N
<i>C. elegans</i>	10.5 (2.1)	5 (1)	2.1	100%	N
<i>S. pombe</i>	> 57a (> 19)	45 (15)	> 1.1	0%	Y

La comparaison du nombre de cassures double-brin, des crossing-overs, et de l'utilisation des cassures double-brin pour l'appariement et la proportion de crossing-overs interférents dans différents organismes. n.d. : pas de données. Les crossing-overs chez la drosophile sont sujets à l'interférence, mais l'existence d'un autre type de crossing-overs n'a pas été décrite.

Récemment, une nouvelle protéine (*TopoVIB-like*) essentielle pour l'activité catalytique qui interagit directement avec *SPO11* et aide à la formation de l'hétéro-dimère des deux orthologues de *SPO11* (*SPO11-1* *SPO11-2*) chez *Arabidopsis*, a été identifiée chez les animaux et les plantes. Des études montrent ainsi que le mode d'action du complexe enzymatique qui procède aux cassures double-brin de l'ADN est conservé et pourrait dériver d'un système ancestral commun de topo-isomérase, même si d'importantes différences structurales entre espèces sont suspectées [38,41].

Les cassures double-brin de l'ADN sont effectuées après la réplication de l'ADN. Cependant, le moment précis varie entre les organismes et va dépendre de la capacité du complexe synaptonémal à se former avec ou sans cassures double-brin ainsi que selon les interactions sous-jacentes entre chromosomes. Chez la levure ou les mammifères, la formation du complexe synaptonémal est dépendante des cassures double-brin, qui se produisent très tôt en leptotène. Alors que chez la drosophile, elles sont toujours formées lors de la prophase, mais après la formation complète du complexe synaptonémal en pachytène [45].

Le nombre de cassures double-brin ainsi que leur localisation sont très finement contrôlés pour éviter tout dommage délétère à la cellule. Un exemple du contrôle exercé est l'induction très forte, après les autosomes, des cassures double-brin au sein des régions pseudo-autosomales des chromosomes sexuels chez les mammifères. Leur ségrégation va dépendre des capacités d'appariement entre les chromosomes X et Y, et des liens physiques créés dans ces régions pseudo-autosomales assurés par les crossing-overs [45]. Chez *Arabidopsis* les cassures double-brin n'apparaissent pas non plus de manière aléatoire sur le génome mais sont présentes en grand nombre (environ 50 par chromosome) [44] (Tableau 1).



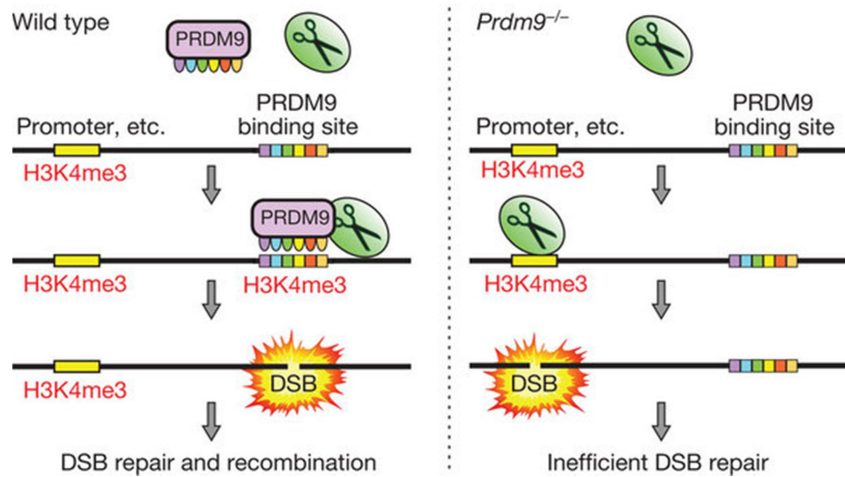


Figure 15 : Modèle pour le rôle de *PRDM9* (d'après [46])

A gauche, les cellules contenant une copie fonctionnelle de *PRDM9*, la machinerie pour les cassures double-brin (ciseaux) est dirigée préférentiellement vers les sites de liaison *PRDM9*. A droite, en l'absence de *PRDM9*, la machine des cassures double-brin se positionne de manière opportuniste à des sites où les marqueurs H3K4me3 sont présents (comme les promoteurs). Ce qui produit une cassure double-brin non réparable et un arrêt de la méiose.

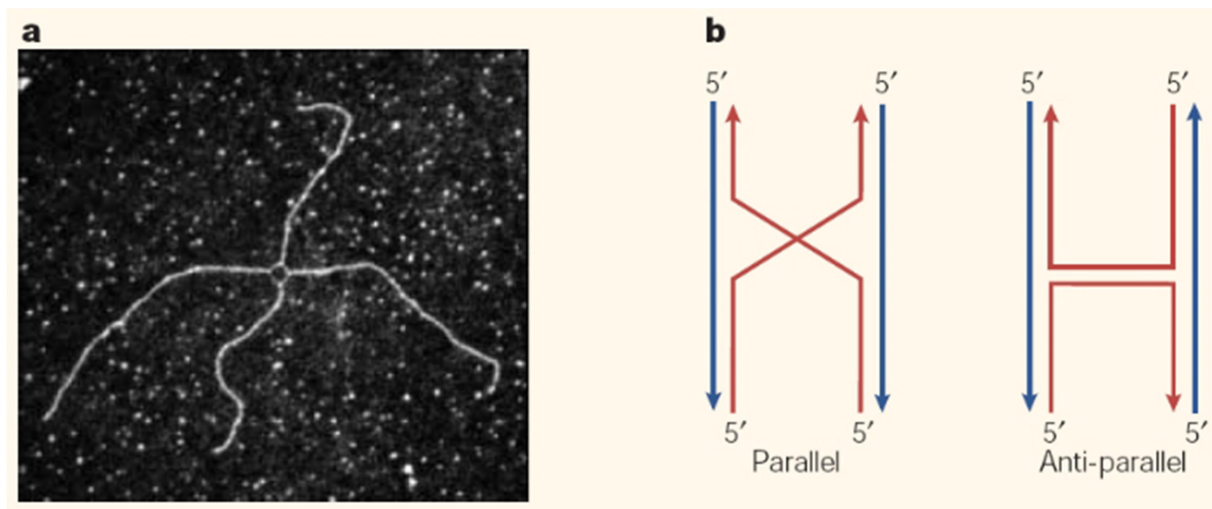


Figure 16 : Structure de la jonction de Holliday (d'après [47])

(a) Image de microscopie électronique d'un intermédiaire de recombinaison. Dans cette image la jonction de Holliday a été partiellement dénaturée pour aider sa visualisation. (b) Deux configurations possibles pour la jonction de Holliday, avec l'ADN figurant dans le sens parallèle (à gauche) ou la configuration antiparallèle (à droite).

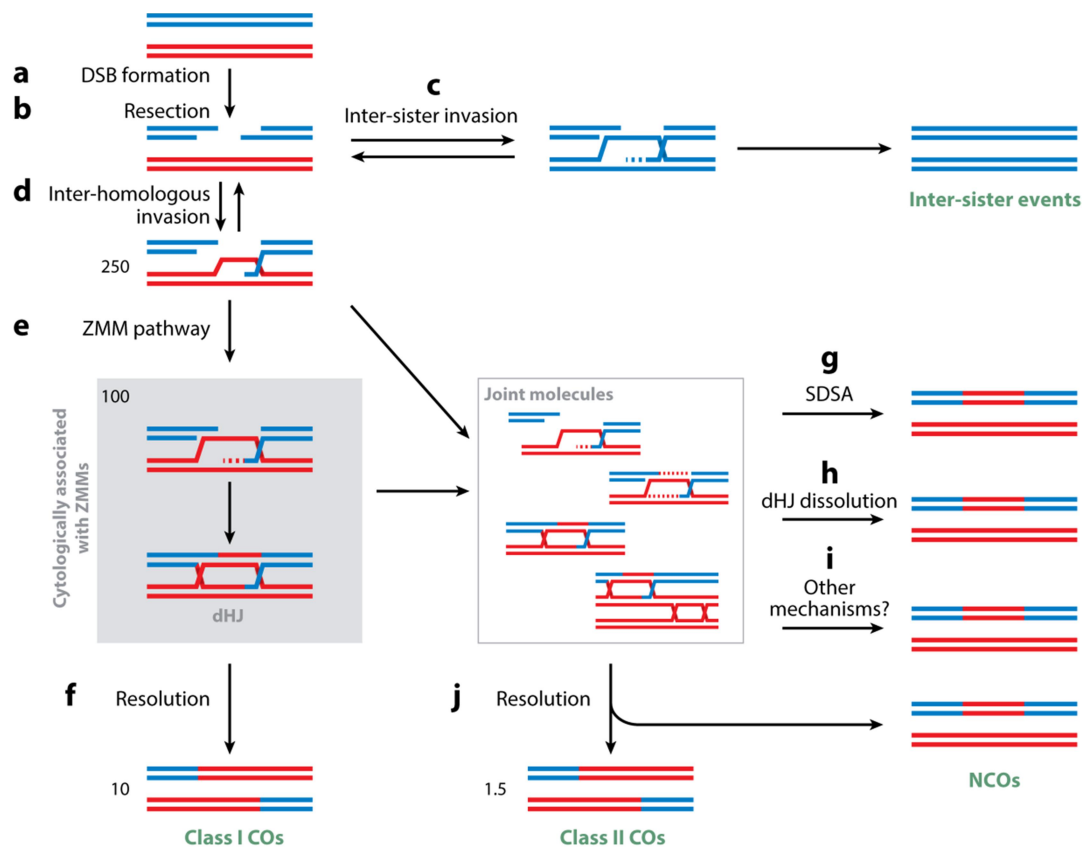
## 5. Localisation des Cassures

Les cassures double-brin sont générées à l'intérieur des boucles de chromatine. Ces cassures sont produites et associées physiquement avec les axes qui formeront le complexe synaptonémal. Elles se situent principalement dans des régions de chromatine ouverte et leur localisation dépend de plusieurs facteurs : modification chromatinienne, cohésion, composition en bases et structure chromosomique. Ainsi, elles se produisent dans des régions préférentielles que l'on nommera « point chaud » (ou hotspot) [45].

Chez la souris et l'humain, un facteur supplémentaire est présent : PRDM9 (Figure 15). Cette protéine exprimée spécifiquement dans les méiocytes est une histone méthyltransférase qui contient un domaine en doigt de zinc permettant une fixation à l'ADN et un domaine méthyltransférase PR/SET qui est capable de catalyser la triméthylation de la lysine 4 de l'histone H3 (H3K4me3). Aussi *PRDM9* va influencer la localisation de la recombinaison par liaison à des sites spécifiques de l'ADN et par la H3K4-triméthylation dans le voisinage des nucléosomes. Finalement, ce signal va stimuler la recombinaison homologue par la création de cassures double-brin dans ces régions spécifiques [48,49].

## 6. Résolutions des cassures

Différents modèles pour la réparation des cassures double-brin ont été proposés depuis la caractérisation d'une structure essentielle, la jonction de Holliday. Cette structure nommée d'après Robin Holliday a été identifiée en 1964 comme une structure particulière d'échange d'information génétique chez les champignons. La jonction de Holliday est une structure ressemblant à une croix qui possède 4 doubles hélices d'ADN attachées de manière covalente les unes aux autres par des échanges mutuels de portions de brins d'ADN [50] (Figure 16). Cette structure identifiée et étudiée chez la levure est présumée être le précurseur universel de la recombinaison [51].



**AR** Mercier R, et al. 2015.  
 Annu. Rev. Plant Biol. 66:297–327

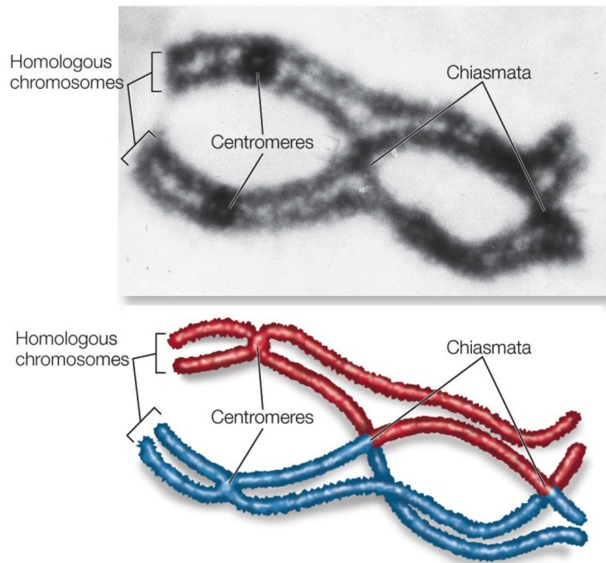
Figure 17 : **Modèle mécanistique de la recombinaison méiotique**(d’après [16])

La recombinaison méiotique est initiée par la formation d'un grand nombre de cassures double-brin (DSB) (a) qui sont traitées (b) pour donner des extrémités 3'-OH d'ADN simple-brin. Cet ADN peut ensuite envahir, soit la chromatide sœur intacte (c) ou l'une des deux chromatides homologues, formant ainsi une boucle (D loop) (d). Les intermédiaires inter-homologues peuvent être protégés par des composants de la voie de ZMM (e), créant une double jonction de Holliday (dhj) qui peuvent être résolus par des crossing-overs (CO) de classe I (f). De manière alternative, les intermédiaires peuvent être résolus en non-crossing-overs (NCO) par le biais de différents mécanismes, y-compris la synthèse d'ADN dépendante du recrutement d'un brin (SDSA) (g), la dissolution de la double jonction de Holliday (h), et éventuellement d'autres mécanismes (i). En outre, une voie indépendante des ZMM produit des CO de classe II non soumis à l'interférence (j). Le nombre estimé de chaque intermédiaire chez *Arabidopsis thaliana* est indiqué.

Ce premier modèle a ensuite été actualisé en 1983 par Szostak [52] pour devenir le modèle de cassure double-brin (DSBR). Ensuite, différentes voies ont été ajoutées à ce dernier, afin de prendre en compte les événements présents lors des divisions mitotiques d'après des observations chez la levure : SDSA pour synthesis dependant strand annealing, SSA pour simple strand annealing et le modèle de réplication induite par la recombinaison (BIR) [53].

Aujourd'hui le modèle décrit (Figure 17) pour la recombinaison méiotique explique qu'à la suite de la formation d'une cassure double-brin il y a résection des extrémités pour obtenir une extrémité 3' simple brin. Ensuite, la réparation pourra s'effectuer par invasion de l'une des extrémités 3' simple-brin dans la chromatide sœur ou dans le chromosome homologue intact ce qui va conduire à la formation d'une D-loop (boucle) et d'une jonction de Holliday. La seconde extrémité va s'apparier à la boucle et les portions d'ADN manquantes discontinues seront réparées par synthèse d'ADN. A ce stade les molécules d'ADN peuvent être protégées par des composantes de la voie ZMM des crossing-overs ce qui va permettre la formation d'une double jonction de Holliday qui pourra être résolue en crossing-overs de type I. Sinon, les intermédiaires pourront être résolus en crossing-overs de type II par une voie ZMM indépendante ou en non crossing-overs (NCO) à travers différents mécanismes comme la voie SDSA. Dans ce dernier cas, la boucle se dissocie de l'extrémité envahissante et retourne s'hybrider avec son brin complémentaire d'origine. Il y a ensuite synthèse d'ADN au niveau de la cassure d'origine puis ligation [16].

Le grand nombre de cassures double-brin par rapport au nombre de crossing-overs réalisés (25-30 cassures pour 1 crossing-over chez *Arabidopsis*) démontre qu'en plus du contrôle du nombre et de la localisation des cassures double-brin, leur réparation en crossing-over est également très régulée [44].



LIFE 8e, Figure 9.17

LIFE: THE SCIENCE OF BIOLOGY, Eighth Edition © 2007 Sinauer Associates, Inc. and W. H. Freeman & Co.

Figure 18 : Photo et représentation d'un crossing-over (d'après [54])

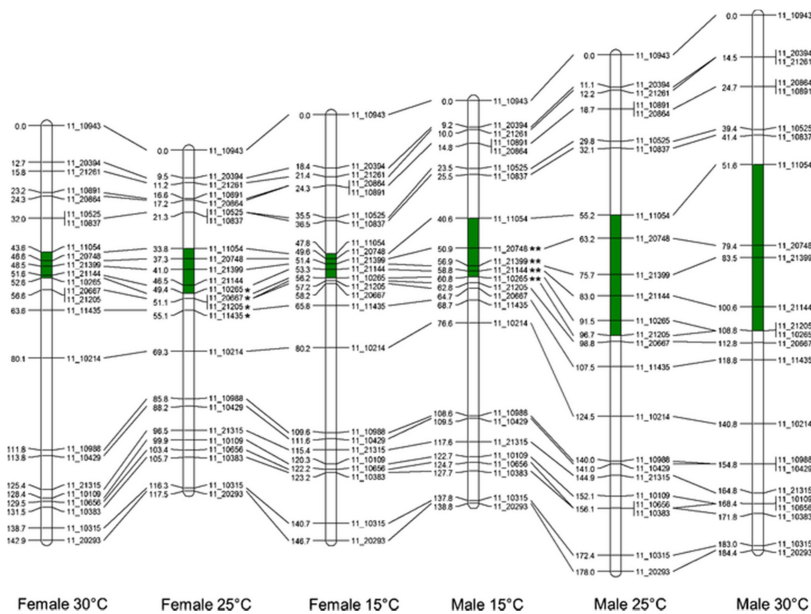


Figure 19 : Comparaison de cartes génétiques du chromosome 2H de l'orge (d'après [55])

Comparaison des cartes génétiques du chromosome 2 H de l'orge obtenue à partir de six populations de backcross 1. Celles-ci ont été obtenues par le croisement entre les variétés Steptoe et Morex en utilisant le parent F1 mâle ou femelle croisé en retour avec le parent Morex. Les plantes ont été cultivées à trois températures différentes. L'intervalle ombré met en évidence la région génétique proximale autour du centromère. Les locus polymorphes qui présentent une distorsion de ségrégation significative sont marqués d'un astérisque (\*,  $P \leq 0,01$ ; \*\*,  $P \leq 0,005$ ).

## II. La recombinaison et le déséquilibre de liaison

---

La transmission des caractères dans une descendance a été décrite en 1865 par Gregor Mendel qui établit la théorie de l'hérédité et de la ségrégation des caractères [56]. Par la suite William Bateson, Rebecca Saunders et Reginald Punnett vont décrire le phénomène de liaison génétique en 1905 qui sera conceptualisé par Thomas Morgan en 1911. Enfin, le concept de crossing-over sera développé en 1909 par Frans Janssens et confirmé chez le maïs en 1931 par Harriet Creighton et Barbara McClintock [57]. Ce phénomène de crossing-over (Figure 18) a lieu lors de la méiose et permet le brassage génétique intra-chromosomique et la génération d'une nouvelle combinaison d'allèles à l'échelle du chromosome. Néanmoins, ce phénomène est rare et n'est pas présent partout de manière homogène sur les chromosomes [39].

En effet, la formation des crossing-overs est très contrôlée et un minimum de un crossing-over par bivalent est observé (crossing-over obligatoire). Ainsi le nombre de crossing-overs excède rarement trois par bivalents dans des conditions normales. Par ailleurs lors de la réduction du nombre de cassures double-brin le nombre de crossing-overs est maintenu. Ce phénomène est appelé homéostasie des crossing-overs [16,58,59]. D'autres phénomènes comme l'hétérochiasmie, qui est la variation du taux de recombinaison en fonction du sexe, ou encore des variations du taux de recombinaison en fonction de la température, ont été montrés chez *Arabidopsis* et chez l'orge [55,60] (Figure 19). De plus des variations selon différents types de stress sont suspectées [61]. Comme décrit dans le modèle mécanistique de la résolution des cassures, il existe deux types de crossing-overs : les CO de Type I, interférents, et les CO de type 2, non interférents. L'interférence a été décrite par Muller en 1916 comme suit : « l'apparition d'un crossing-over interfère avec l'apparition coïncidente d'un autre crossing-over dans la même paire de chromosomes ».



Cette interférence entre crossing-overs a été confirmée chez beaucoup d'espèces et peut être estimée de manière statistique ou via la visualisation de la distribution des nodules de recombinaison ou chiasmas de manière cytologique. Chez Arabidopsis et la levure, les observations conduisent à penser que la grande majorité des crossing-overs est sujet à l'interférence [39]. Ce phénomène d'interférence implique l'existence d'une forme de communication le long des chromosomes. Depuis, plusieurs modèles ont été décrits pour essayer de détailler le fonctionnement de l'interférence : par signal ou par redistribution du stress mécanique... Néanmoins, le mécanisme reste inconnu aujourd'hui.

Les phénomènes de non crossing-overs (NCO) correspondent au processus de réparation de la cassure double-brin lors de la méiose n'étant pas associé à un crossing-over. Dans le cas où ceux-ci sont génétiquement détectables, ils seront qualifiés de conversion génique simple. Leur détection dépendra de la présence ou absence d'un polymorphisme dans la zone de conversion. Chez les plantes les NCO sont très peu documentés car difficile à détecter. Néanmoins, chez Arabidopsis quelques NCO ont été détectés suggérant que ces événements existent mais sont rares. Cependant ces résultats peuvent être dus à l'incapacité de les détecter par manque de polymorphisme [16], ou à une réparation des cassures double-brin en utilisant la chromatide sœur comme observé chez la levure.

## **1. Méthodes de détection**

De manière classique les crossing-overs sont détectés par l'utilisation de marqueurs génétiques liés physiquement qui vont montrer une transition d'un parent vers l'autre le long d'une distance physique dans une descendance. Chez les champignons l'observation de tétrades permet de dénombrer le nombre de crossing-overs entre deux marqueurs connus pour régir la couleur des gamètes [62]. Cette approche a récemment été étendue à la souris et aux plantes.



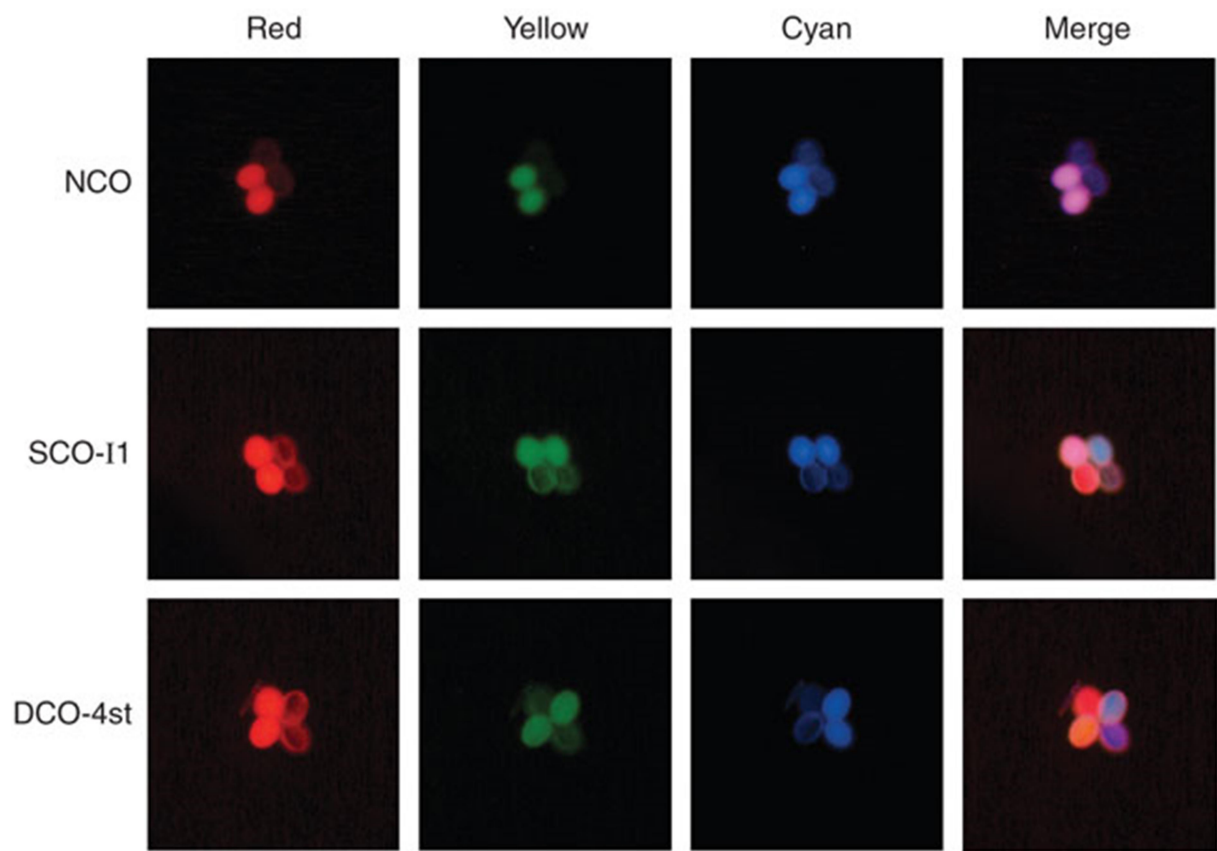


Figure 20 : Visualisation des tétrades fluorescentes d'Arabidopsis (d'après [62])

Chez *Arabidopsis*, l'insertion de constructions génétiques va permettre la synthèse de deux molécules fluorescentes au sein des spores chez des individus sauvages et le mutant quartet, ce dernier permettant aux quatre cellules haploïdes de la méiose mâle de rester en tétrade [63,64] (Figure 20). Les différentes couleurs permettent de visualiser directement le nombre de crossing-overs et de non crossing-overs entre les deux marqueurs fluorescents. Néanmoins cette technique n'est pas encore disponible chez tous les organismes modèles et implique la génération d'OGM spécifiques.

Du fait de la chute des coûts de séquençage, une autre approche consiste à séquencer des descendances afin de détecter les CO présents dans une population. Cependant, du fait du nombre limité de CO par génération ces démarches restent difficiles à mettre en œuvre car il est fastidieux de générer de larges populations. Pour contourner ces difficultés de nouvelles stratégies ont vu le jour en utilisant le typage des gamètes mâles (Pollen ou Sperm Typing) [63]. Récemment, l'avancée des techniques d'amplification génome entier couplées au tri de pollen ont permis la mesure et l'identification de crossing-overs à plusieurs locus chez l'orge [65]. Enfin d'autres méthodes comme la chromatine immuno-précipitation (ChIP-Seq) de protéines dirigées vers des protéines impliquées dans la recombinaison peuvent être employées [63].

Une approche de cartographie historique peut également être utilisée. Cette méthode sera détaillée dans la partie traitant le déséquilibre de liaison.

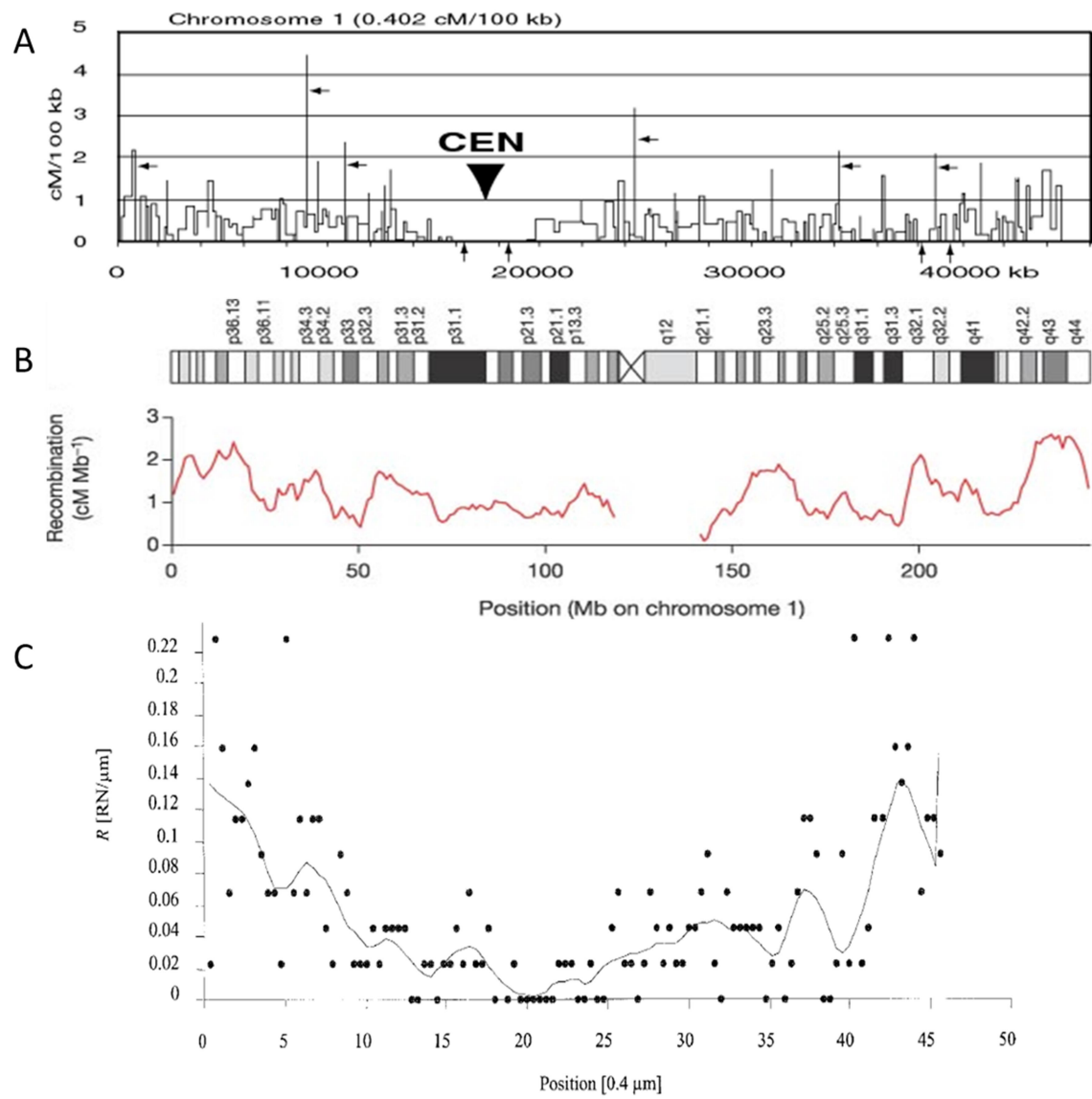


Figure 21 : Exemples de Pattern de Recombinaison (d'après [66–68])

(A) Le riz, (B) l'homme, (C) le maïs.

## 2. Répartition et facteurs de variation des CO

Les crossing-overs vont se distribuer dans deux types de zones le long des chromosomes [39] : des zones chaudes ou zones recombinogènes présentant un taux de CO au-dessus de la moyenne et des zones froides avec un taux de CO en-dessous de la moyenne. Chez toutes les espèces de plantes, la distribution des événements de crossing-over est non homogène (Figure 21). Chez, l'*Allium fistulosum* les crossing-overs semblent se situer près du centromère ; le riz, la tomate et *Arabidopsis* présentent un taux de recombinaison variable le long du chromosome. Finalement, chez les espèces à gros génomes comme le blé, le maïs ou l'orge, la recombinaison est très favorablement biaisée vers les régions distales des chromosomes [16,39]. Ce pattern pourrait peut-être être s'expliquer par une initiation rapide de la recombinaison dans les parties terminales des chromosomes chez l'orge [26]. Récemment, des études utilisant des segments transloqués de blé suggèrent que le pattern de recombinaison est prédéterminé et indépendant de la localisation distale ou proximale par rapport aux centromères [69,70]. Cependant l'abolition des événements de recombinaison dans les régions centromériques reste une constante chez toutes les espèces. Chez *S. cerevisiae*, *Arabidopsis*, le blé tendre et l'humain, 80% des événements de recombinaison se produisent dans moins d'un quart du génome. Néanmoins, les raisons d'une telle hétérogénéité restent inconnues [16].

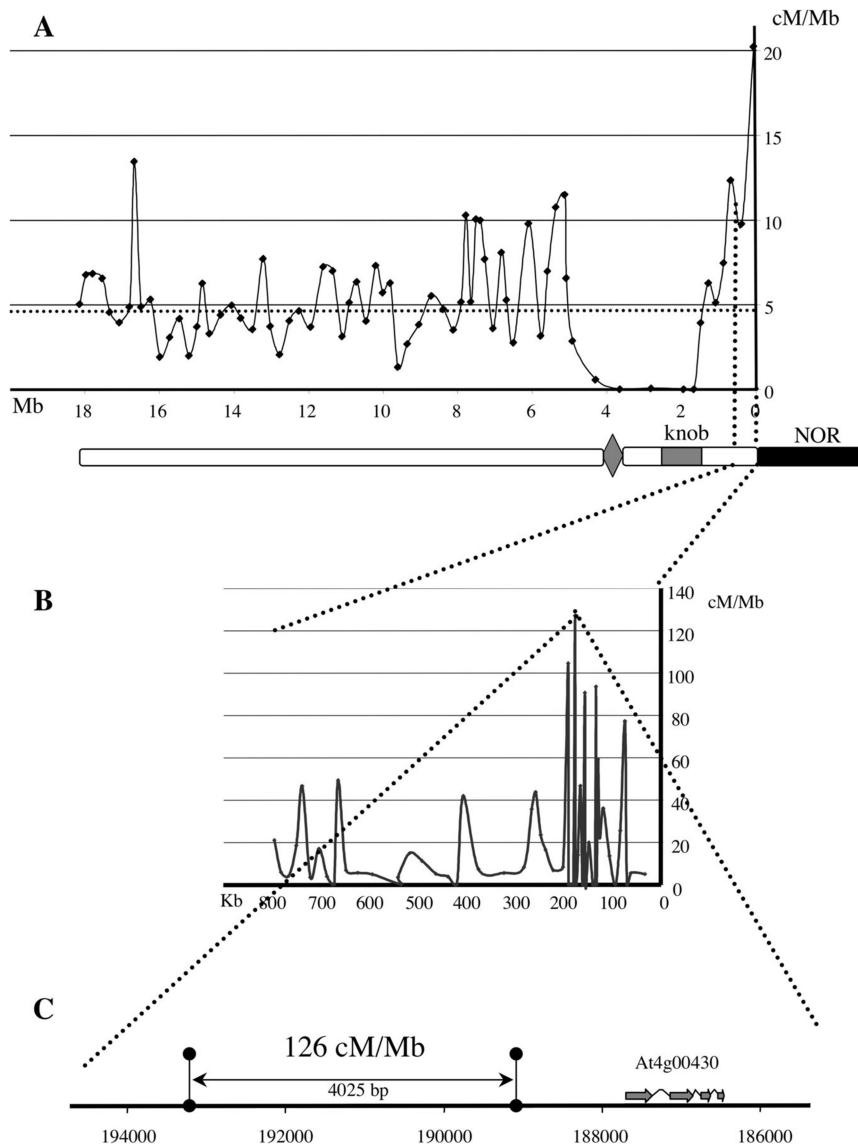


Figure 22 : La Recombinaison le long du chromosome 4 d'*Arabidopsis* (d'après [71])

(A) Variation du taux de crossing-overs le long du chromosome 4 d'*Arabidopsis thaliana*. (B) Distribution des crossing-overs dans les premiers 800kb du début du bras court. Les crossing-overs sont concentrés dans 10 régions de moins de 5kb sur le bras court du chromosome 4, ce qui génère un taux de recombinaison au moins 5 fois plus grand que la moyenne du génome (4.6cM/Mb). (C) Cluster de crossing-overs dans une petite partie d'ADN.

### 3. Point chaud de Recombinaison

La notion de points chauds de recombinaison résulte de l'observation de variations du taux de recombinaison selon les échelles considérées (Figure 22). Un point chaud de recombinaison se définit ainsi comme « une zone du génome (de 1-2kb) présentant un taux de recombinaison élevé » [72]. Elle correspond donc à une localisation physique définie dans laquelle se produisent les évènements.

Le nombre de points chauds étudiés chez les plantes de manière fine reste faible et varie de 3 chez le maïs, 4 chez *Arabidopsis* et le riz et un chez le blé [63,67,73–81]. L'obtention d'une résolution fine dans le voisinage des points chauds est complexe. Cette complexité n'est pas circonscrite aux plantes. Chez l'humain, des études réalisées par séquençage ont mis en évidence que seul 13 à 45% des crossing-overs sont cartographiés dans des intervalles de moins de 30kb. Récemment, chez les oiseaux, 279 crossing-overs ont été localisés dans des intervalles de moins de 10kb [63,82].

### 4. Caractéristiques des points chauds

Malgré les difficultés qui résultent du faible nombre de CO et de la difficulté de les cartographier dans de petites régions, des caractéristiques générales de l'environnement des points chauds ont pu être dégagées. Des études montrent que les points chauds de recombinaison semblent se situer dans les régions géniques et notamment dans leurs parties 5' et 3'. Chez le maïs l'étude du locus *Bronze* a montré une localisation préférentielle des CO dans les gènes, même si la résolution des cartes génétiques ne permet pas de conclure sur la position exacte des points chauds de recombinaison [74,75]. Chez *Arabidopsis* les points chauds de recombinaison étudiés ont notamment été montrés avec un gradient de localisation avec un maximum au niveau des régions 5' des gènes et un minimum au niveau des régions 3'.

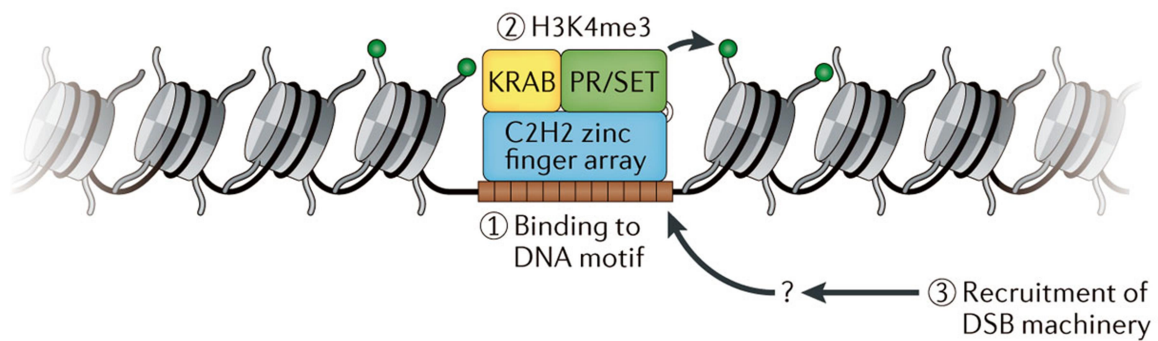


De plus, leur présence a été confirmée dans les régions promotrices et terminatrices des gènes par pollen-typing pour quatre points chauds [77,79,83]. Ces études suggèrent donc que chez les plantes les points chauds de recombinaison se situeraient préférentiellement dans les zones géniques, et plus particulièrement dans les régions 5' peut-être à cause de motifs spécifiques à l'organisation de la chromatine associée à la transcription par la Polymerase II [84]. Des résultats similaires ont été observés chez la souris mutante dont le système PRDM9 est inactivé.

Un autre type d'élément constitutif du génome semble avoir un impact : les éléments transposables. Historiquement la séquence cible de PRDM9 a été trouvée surreprésentée au niveau de THE1A/B, un groupe particulier d'éléments répétés apparentés aux rétrovirus [85–87]. De plus chez le maïs, une étude a montré que les événements de recombinaison et les transposons de type *Mutator* apparaissaient préférentiellement dans les régions 5' des gènes, et qu'ils étaient corrélés positivement avec la présence d'une marque épigénétique (H3K4me3) associée à une chromatine ouverte. Par ailleurs, chez le maïs une réduction de la recombinaison dans les régions très méthylées riches en éléments transposables a été observée [75,88].

Des éléments transposables ont été trouvés chez toutes les espèces chez qui ils ont été recherchés. Cependant leur abondance, diversité et activité restent très variables selon les espèces. De même, les mécanismes qu'ils ont développés afin de se maintenir sont tout aussi divers [89–92]. Dans certains cas l'insertion se fait préférentiellement près d'un gène peut-être à cause de la conformation ouverte de l'ADN dans ces régions. Cela pose la question du rôle des marques épigénétiques car la conformation de l'ADN, sous contrainte de celles-ci, peut avoir une grande importance pour prévenir la recombinaison dans les éléments transposables et éviter de la recombinaison ectopique [93].





Nature Reviews | **Genetics**

Figure 23 : **Modèle expliquant le rôle de PRDM9 dans la localisation des cassures double brin (d'après [48])**

Le domaine de liaison PR/SET 9 (*PRDM9*) se lie à un motif spécifique de l'ADN (carrés marron) à l'aide de la structure en doigts de zinc C2H2 (rectangle bleu). Par la suite, le domaine PR/SET (région homologue de PRDI-BF1 et RIZ, qui est une sous-classe des domaines SET ; rectangle vert) favorise la triméthylation de la lysine 4 de l'histone H3 (H3K4me3, sphères vertes) sur les nucléosomes adjacents. Le domaine 'Krüppel-associated box' (KRAB) pourrait avoir des interactions avec d'autres protéines. Ces étapes et peut-être des événements supplémentaires permettent le recrutement de la machinerie des cassures double brin, dont *SPO11* à l'aide d'un mécanisme inconnu.

Dans le système PRDM9 (Figure 23), les marques épigénétiques sont d'une grande importance. En effet, ce système utilise un motif spécifique CCNCCNTNNCCNC pour venir déposer des marques H3K4me3. Les recherches chez les plantes n'ont pas permis de mettre en évidence de tels motifs et le système PRDM9 n'existe pas [94]. Cependant quelques caractéristiques de séquences ont pu être montrées telles qu'une présence accrue des répétitions de motifs CCT ou CCN ainsi que des régions riches en A [77,95]. Ces dernières régions ont été montrées comme ayant une influence sur l'exclusion des nucléosomes créant ainsi des régions plus ouvertes [95,96]. De plus, les régions hypo-méthylées ou portant certaines marques (H3K4me3, H2a.z) épigénétiques particulières ont été montrées associées avec la recombinaison chez *Arabidopsis* et l'orge [97,98].

## 5. Evolution des points chauds

Différentes observations conduisent à penser que la localisation des points chauds de recombinaison n'est pas là même au cours du temps. Chez la souris, il a été montré des positions distinctes entre différents génotypes mettant en évidence une évolution très rapide des points chauds. A une grande échelle de temps, comme entre l'homme et le chimpanzé qui partagent pourtant plus de 99% [99] de leurs séquences génétiques, la position des points chauds n'est pas conservée. Sur un pas de temps plus court, comme au sein du genre Humain, les comparaisons entre cartographies historiques basées sur le déséquilibre de liaison (DL) et la recombinaison actuelle identifiée sur des gamètes, montrent qu'il n'y a pas forcément de recouvrement entre les deux. Ceci témoigne ainsi d'une évolution très rapide des points chauds chez l'homme [48]. Le système PRDM9 serait potentiellement responsable de ce changement rapide de position. [100]. Tout d'abord, les analyses visant à estimer le taux de mutation montrent que la partie doigt de zinc de *PRDM9* qui vient se fixer à l'ADN présente un taux de mutation très élevé ce qui pourrait expliquer des localisations différentes dues à la diversité allélique [49].

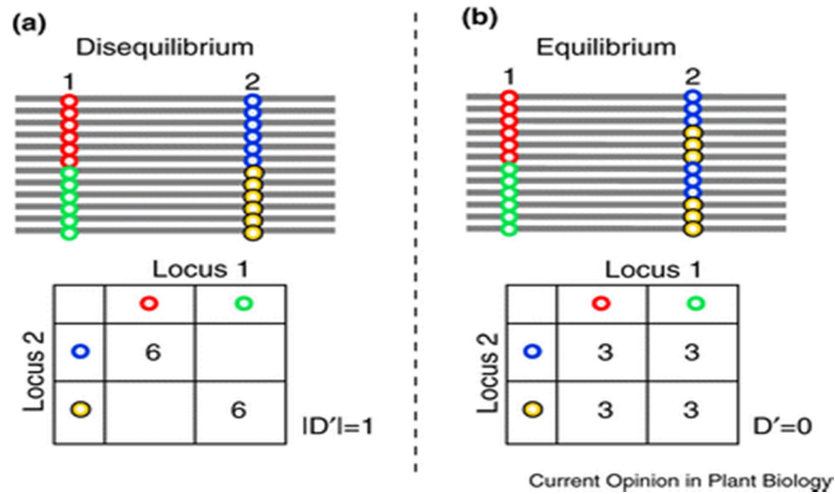


Figure 24 : Représentation schématique du déséquilibre (a) et (b) équilibre de liaison entre deux locus [101]

(a) Lorsque le déséquilibre de liaison est total, tous les individus qui possèdent l'allèle rouge au locus 1 ont nécessairement l'allèle bleu au locus 2. (b) Quand il y a équilibre de liaison, l'individu qui possède l'allèle rouge au locus 1 à autant de chance de posséder l'allèle jaune ou bleu au locus 2. Les tableaux de contingence respectifs sont fournis et les valeurs du déséquilibre ( $D'$ ) sont données

Ensuite le mécanisme de cassure double-brin et de réparation pourrait être lui-même responsable en provoquant la modification par conversion génique des séquences qui vont présenter le plus souvent des cassures [86,100]. Chez les plantes, le système PRDM9 n'a jamais été présent ou a été perdu [94]. Seules deux espèces ont fait l'objet de cartographies historiques : *Arabidopsis* [63] et *Medicago truncatula* [102]. Dans le cas d'*Arabidopsis*, les signaux de recombinaison ancestrale semblent correspondre à la recombinaison actuelle. Concernant *Medicago truncatula* la question n'a pas été traitée.

## 6. Déséquilibre de liaison

Le déséquilibre de liaison (DL) représente l'association non-aléatoire de deux ou plusieurs locus et résulte d'un déséquilibre gamétique [103]. Néanmoins, la liaison génétique due à la proximité physique des locus est souvent confondue dans le déséquilibre de liaison [103].

Le DL est calculé sur une échelle de 0 à 1. La valeur 1 représente un déséquilibre total par rapport à l'association aléatoire, c'est-à-dire une association parfaite entre les allèles. Réciproquement, la valeur 0 représente une ségrégation complètement aléatoire, c'est-à-dire une dissociation complète entre les allèles [103] (Figure 24).

Différents facteurs peuvent impacter le DL, notamment la recombinaison ou la mutation par la création de nouveaux allèles. Plus le taux de recombinaison sera élevé plus il abaissera le DL. Au contraire, la sélection favorisera l'association non aléatoire d'allèles et augmentera ainsi le déséquilibre de liaison d'un point de vue local ou à longue distance [84,103,104].

Aussi, la chute du déséquilibre de liaison sur une très courte distance peut également coïncider avec des événements de non crossing-overs comme cela a été montré chez le blé [76].

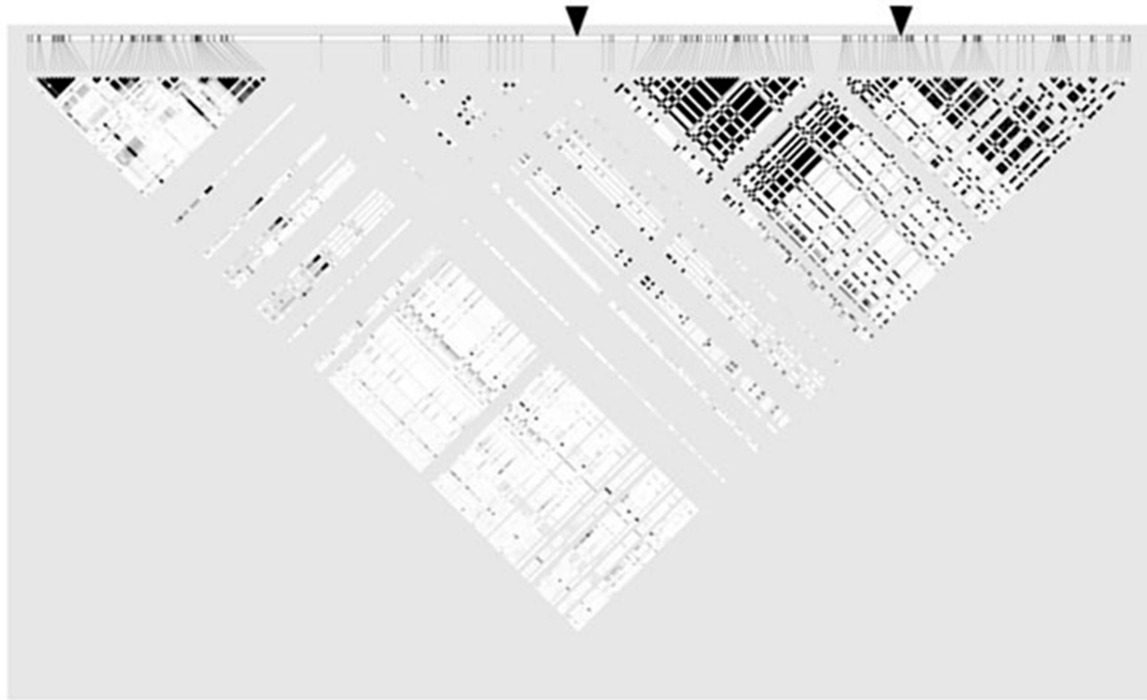


Figure 25 : **Exemple de pattern de DL (Arabidopsis chr 4) (d'après [83])**

Les carrés noirs correspondent à un fort déséquilibre de liaison ( $r^2=1$ ) et les carrés blancs à un déséquilibre faible ( $r^2=0$ ). La localisation de deux potentiels points chauds de recombinaison est indiquée par les triangles noirs.

De cette liaison entre les allèles découle un autre concept : l'autostop génétique. Si des caractères sont liés physiquement, leur ségrégation a plus de chance de se faire ensemble (liaison génétique). Ainsi, un gène sélectionné sera probablement accompagné d'autres gènes localisés dans son voisinage à cause (1) du faible nombre de crossing-overs ; (2) de la répartition non-aléatoire et non-homogène des crossing-overs.

Le déséquilibre de liaison est représenté sous forme de blocs de liaison (Figure 25) qu'il y a un grand intérêt à casser (notamment dans les régions centromériques) dans le cas d'un gène d'intérêt co-ségrégeant avec un gène délétère. Le contrôle de (activer/désactiver) la recombinaison pour casser ces blocs est donc une voie de recherche étudiée actuellement avec grand intérêt. De plus, l'impact de la recombinaison sur l'association non aléatoire des mutations et l'analyse du polymorphisme génétique actuel permet l'étude de la recombinaison ancestrale. En effet, les crossing-overs vont provoquer des chutes du déséquilibre de liaison entre polymorphismes liés qui peuvent être analysées sous contrainte de coalescence [63].

Ce modèle théorique a pour principe le partage entre deux individus d'un ancêtre commun reconstruit au travers d'algorithmes statistiques pour établir les zones de variation qui pourront être expliquées par la recombinaison. Par exemple, si l'on considère deux locus liés génétiquement portant chacun une mutation indépendante a et b, l'obtention de l'haplotype a-b découle soit d'une mutation récurrente de ces locus, soit de la recombinaison qui va associer a et b sur le même chromosome [63] (Figure 26).

Ces approches sont très puissantes puisqu'elles prennent en compte la multitude de méioses qui ont eu lieu au cours de l'histoire des individus. Cependant, la recombinaison n'est pas la seule force génétique pouvant influencer la structure de l'haplotype. Aussi certains auteurs suggèrent de combiner ces approches de cartographie ancestrale avec la cartographie génétique directe [63].

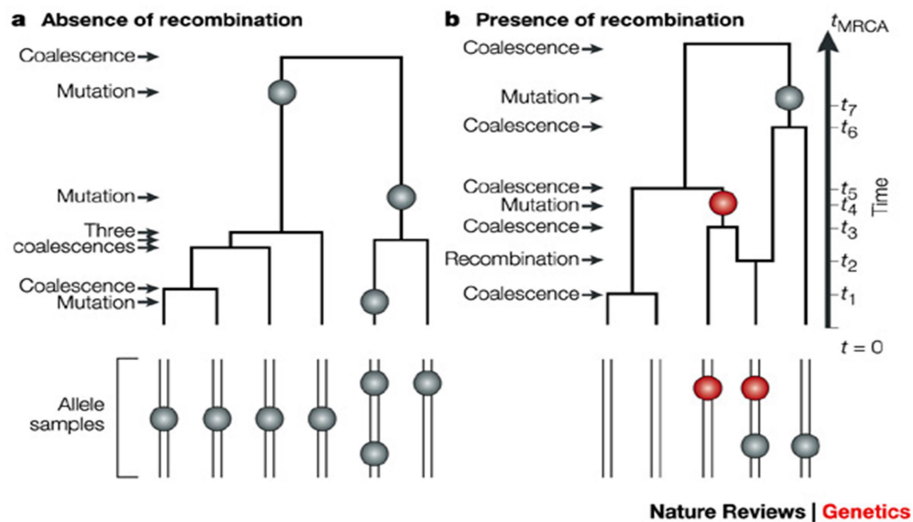


Figure 26 : Simulation de coalescence (d'après [105])

La coalescence est une description statistico-généalogique d'un échantillon d'allèles qui est utile pour déterminer les paramètres populationnels comme l'effet de la recombinaison. Une simulation de la coalescence peut être effectuée en permettant à des paires de lignées de 's'agréger' en remontant dans le temps (panel a). Le point dans le temps où tous les allèles ont fusionné est connu comme « le temps moyen du plus récent ancêtre commun » ( $t_{MRCA}$ , time of the most recent common ancestor). La vitesse à laquelle des échantillons ou des branches fusionnent dans la généalogie dépend des tailles de l'échantillon et de la population. Plus il y a d'individus échantillonnés, plus le regroupement des lignées sera rapide. Par la suite, les mutations neutres peuvent être placées de façon aléatoire sur les branches de la généalogie à une vitesse constante  $\Theta/2$  ( $\Theta$ =taux de mutation ; pour les haploïdes,  $\Theta/2 = 2*Ne*m$ , où  $Ne$  est la population efficace et  $m$  est le taux de mutation par génération par locus). Les cercles sur les branches représentent des mutations. Si les mutations se produisent à une vitesse constante, alors les généalogies anciennes seront censées avoir un plus grand nombre de mutations, et le  $t_{MRCA}$  de tous les allèles, échantillonnés à partir d'une seule et même population, sera directement lié au niveau de polymorphisme observé dans la population. Cela montre comment les effets de la généalogie et l'histoire des mutations sont conservés de manière séparée. De même, le nombre d'événements de recombinaison qui est observé dans une population dépend également du temps du  $t_{MRCA}$  (b). Plus la population sera ancienne, plus la probabilité que des événements de recombinaison se produisent entre les individus ou dans l'histoire de l'échantillon est grande. Les effets de recombinaison sur l'histoire de l'échantillon sont donc fonction non pas du taux de recombinaison absolue, mais du produit du taux de recombinaison par gène par génération de crossing-over ( $r$ , longueur de la carte génétique) par la taille effective de la population ( $Ne$ ). Le schéma b montre comment la recombinaison brise les relations de liaison; avant que l'événement de recombinaison n'ait eu lieu, les mutations (gris et cercles rouges) étaient sur différentes lignées. Cependant, après l'événement, elles se retrouvent sur le même haplotype. Comme montré en dessous de la figure de la généalogie, les différentes parties d'un allèle ont différentes relations ancestrales. Les temps pour chaque agrégation et événement de recombinaison sont présentés. En utilisant le modèle de coalescence, la probabilité d'apparition d'un événement de recombinaison peut être modélisé.

Plusieurs indices suggèrent que les blocs de liaison peuvent être communs entre plantes ce qui montrerait un possible mécanisme de maintien [104]. Chez les plantes la structure du DL produite par l'évolution des points chauds de recombinaison n'a pas encore été démontrée, contrairement à l'humain où sa chute brutale interrompant un bloc de liaison dans une région reflète les effets historiques de la recombinaison dans la région. Enfin, chez *Saccharomyces paradoxus*, l'étude des points chauds de recombinaison basée sur le DL a montré que leur activité est conservée chez les différentes souches [106].



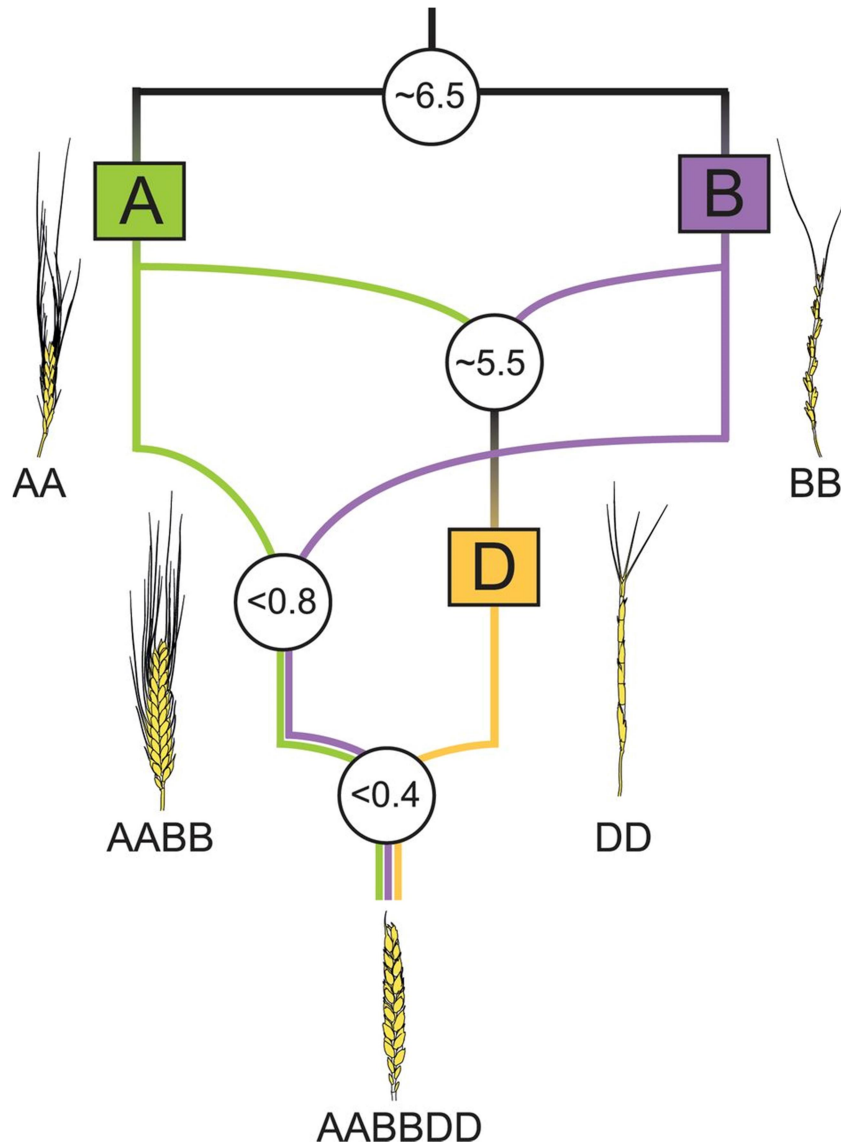


Figure 27 : La formation du blé tendre (d'après [107])

Les dates approximatives pour la divergence et les trois événements d'hybridation sont données dans des cercles blancs en millions d'années. La différenciation des lignées de blé (*Triticum* et *Aegilops*) à partir d'un ancêtre commun avec les génomes A (vert) et B (violet) a commencé il y a ~ 6,5 Ma. La première hybridation a eu lieu il y a ~ 5,5 Ma entre des espèces portant les génomes A et B et a conduit à l'ancêtre de la lignée du génome D. La deuxième hybridation, entre un parent proche d'*Ae. speltoides* (SS, proche du génome BB) et *T. urartu* (AA), a donné naissance au blé amidonnier allo-tétraploïde (*T. turgidum*; AABB) ancêtre du blé dur actuel (*T. durum*). Le blé tendre (*T. aestivum*) s'est formé par hybridation entre le blé amidonnier et *Ae. tauschii* (DD, orange). Les inflorescences (épis) illustrent les phénotypes typiques des espèces existantes étroitement liées à celles qui sont impliquées dans les événements de polyploïdisation.

### III. Le blé tendre

---

#### 1. Généralités

Le blé tendre (*Triticum aestivum* L.) est une plante monocotylédone annuelle de la famille des *Poaceae*. Exploité sur près de 220 millions d'hectares (3 fois la surface de la France) c'est l'une des trois céréales majeures cultivées à travers le monde devant le maïs et le riz (Faostat 2013). En 2013 la production mondiale de blé tendre était d'environ 713 millions de tonnes (38 Mt, juste pour la France, 5<sup>ème</sup> rang mondial). Le blé représentent la principale source calorique et protéique pour l'alimentation humaine avec 19-21% des apports au niveau mondial. Le blé tendre est une espèce allo-hexaploïde (AABBDD) résultant de deux croisements successifs interspécifiques (Figure 27). Le premier a eu lieu il y a environ 0,8 millions d'année ans entre *T. monococcum ssp urartu* (génomme A) et une espèce encore inconnue (génomme B) issue de la section Sitopsis dont l'espèce la plus proche est *A. speltoïdes* (génomme S) pour former *T. turgidum ssp dicoccum* (AB), ancêtre du blé dur actuel (*T. turgidum ssp durum*). Dans un second temps cette espèce tétraploïde s'est hybridée avec *A. tauschii* (génomme D) pour former *T. aestivum* (AABBDD). La datation de cet évènement par horloge moléculaire le situe il y a moins de 0,4 millions d'année [107]. Cependant il est admis que le blé tendre s'est formé il y a  $\approx 10\,000$  avec la naissance de l'agriculture car aucun ancêtre naturel hexaploïde n'a été retrouvé [108]. Il y a un fort degré d'homologie entre les chromosomes issus des génomes A, B et D qui sont pour cela appelés « homéologues ».

Par ailleurs, le génome est de grande taille (17 Gb soit cinq fois le génome humain et 160 fois celui d'*Arabidopsis*) et majoritairement constitué de séquences répétées (80%). Ces différentes caractéristiques ont longtemps retardé les études sur la structure du génome et en particulier sur le profil de recombinaison.

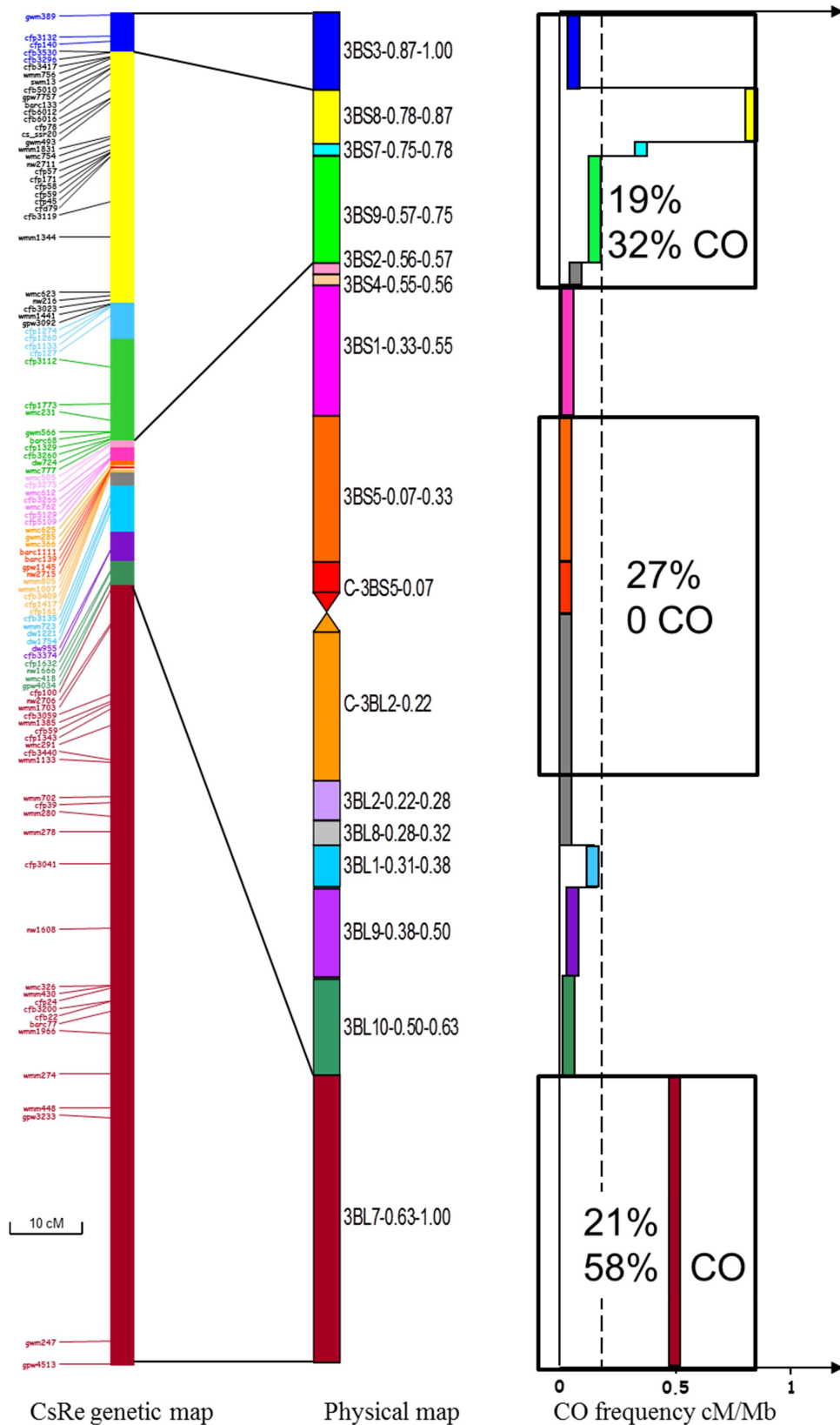


Figure 28 : Distribution de la recombinaison le long de la carte physique du chromosome

3B (d'après [109])

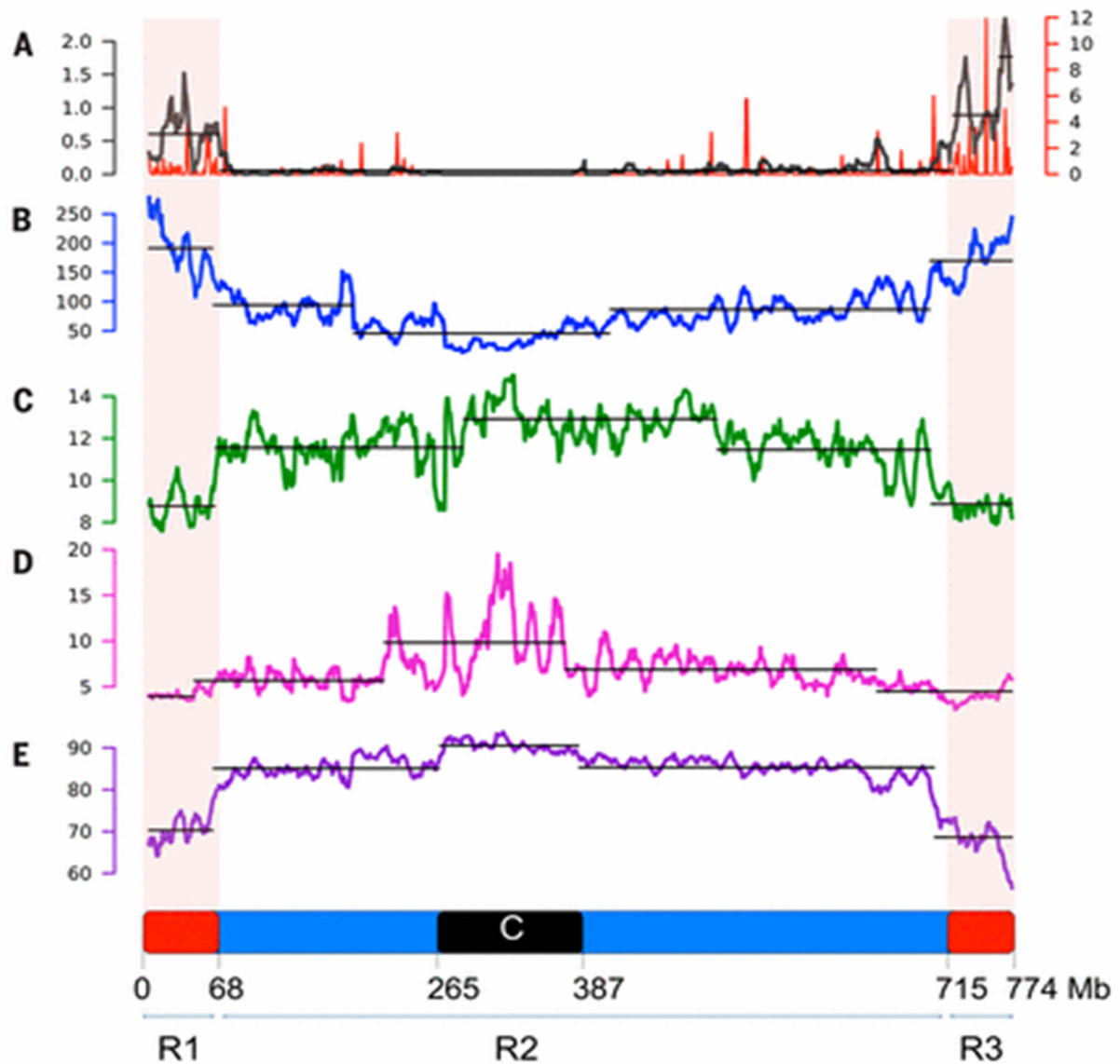
## **2. Analyses préliminaires de la recombinaison chez le blé**

Initialement, la visualisation cytologique des chiasmas en 1993 sur des lignées de blé dur [110] a permis de montrer que la recombinaison se situait dans les régions distales des chromosomes du groupe d'homéologie 3. D'autres études plus exhaustives des différents groupes d'homéologie en utilisant les cartes physiques disponibles ont suggéré une recombinaison liée aux portions riches en gènes [111–114].

De plus, l'utilisation de lignées de délétions ou des lignées de translocation et les analyses par cytologie du nombre de chiasmas [69,115] ont permis de montrer que la formation des chiasmas se fait dans des régions spécifiques et que même la délétion des parties distales ou le retournement du bras de chromosome ne modifie pas la localisation de la recombinaison. Ces résultats suggèrent ainsi que la structure et le contenu de ces régions permet la recombinaison. Le blé tendre semble également présenter de l'interférence ainsi que des différences de taux de crossing-overs entre mâle et femelle (hétérochiasmie) [109,116–118]. Chez le blé il existe aussi une sensibilité à la température pour le déroulement de la méiose. Ainsi une température de 32°C semble affecter l'appariement des chromosomes et générer des univalents [119]. De plus, une étude mentionne la présence de gènes impliqués dans la formation des chiasmas sur le groupe 5 en cas de stress par le froid (10°C) [120].

## **3. Analyse de la recombinaison sur le chromosome 3B**

L'étude de la recombinaison le long du chromosome 3B en parallèle des cartes physiques de délétion [109,110] a permis de montrer que la recombinaison s'effectuait à 90% dans les régions sub-téломériques distales qui ne représentent que 40% du chromosome (Figure 28).



F

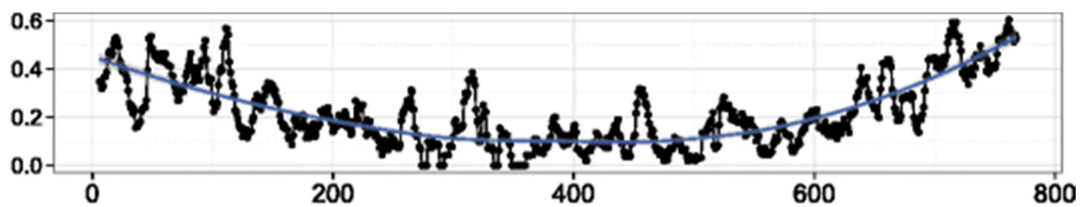


Figure 29 : Le partitionnement du chromosome 3B (d'après [121,122])

A : Taux de recombinaison (cM/Mb fenêtre glissante de 10Mb en noir et 1Mb en rouge) B : Densité en gènes (CDS/10Mb). C : Nombre de conditions. D : Nombre moyen de transcrits alternatifs par gène exprimé. E : Contenu en éléments transposables (%). F : répartition des gènes non synténiques. Les régions distales R1 et R3 sont représentées en rouge. En noir, (C) représente le centromère et les régions péri-centromériques.

Plus récemment, lors de l'ancrage physique de la séquence du chromosome 3B [121], 3075 marqueurs génétiques répartis sur le chromosome ( $38,2 \pm 9,4$  SNPs / 10MB) ont permis de génotyper 284 lignées recombinantes (Single Seed Descent) de la population Chinese Spring X Renan ainsi que 720 individus provenant des ressources génétiques de blé. Ces nouveaux résultats ont permis de confirmer et d'affiner les résultats précédents, et montrent que toute la recombinaison a lieu dans seulement 13% du chromosome [121]. De plus le taux moyen de recombinaison du chromosome entier de 0,16 cM/Mb a pu être calculé.

Ce taux est variable le long du chromosome 3B et permet de le segmenter en 3 parties : les deux zones distales R1 (68 Mb) et R3 (59 Mb) ayant respectivement un taux de recombinaison de 0,60 cM/Mb et de 0,96 cM/Mb alors que la zone R2 où se trouve le centromère représente 649 Mb et n'a un taux de recombinaison que de 0,05cM/Mb [121]. Ces résultats ont permis d'identifier 22 régions, possédant un taux de recombinaison supérieur à 10 fois la moyenne du chromosome, pouvant être considérées comme des points chauds de recombinaison [121].

#### 4. Un partitionnement qui n'est pas limité à la recombinaison

Le chromosome 3B peut être partitionné en trois régions : R1 et R3 qui sont des régions recombinogènes et distales, et la région R2 péracentromérique et centromérique (Figure 29). La densité en gènes, la répartition des gènes non synténiques (la **synténie** est définie comme la conservation de l'ordre des gènes sur un même segment d'ADN entre deux espèces apparentées. Dans ce cas, elle est mesurée par rapport au riz, au maïs, au sorgho et à *Brachypodium*. Les gènes non-synténiques sont ceux qui ne sont pas retrouvés à la place attendue dans les espèces apparentées), le nombre de conditions (combinaison tissu/stade de développement) d'expression et les éléments transposables suivent ce partitionnement [122–124].

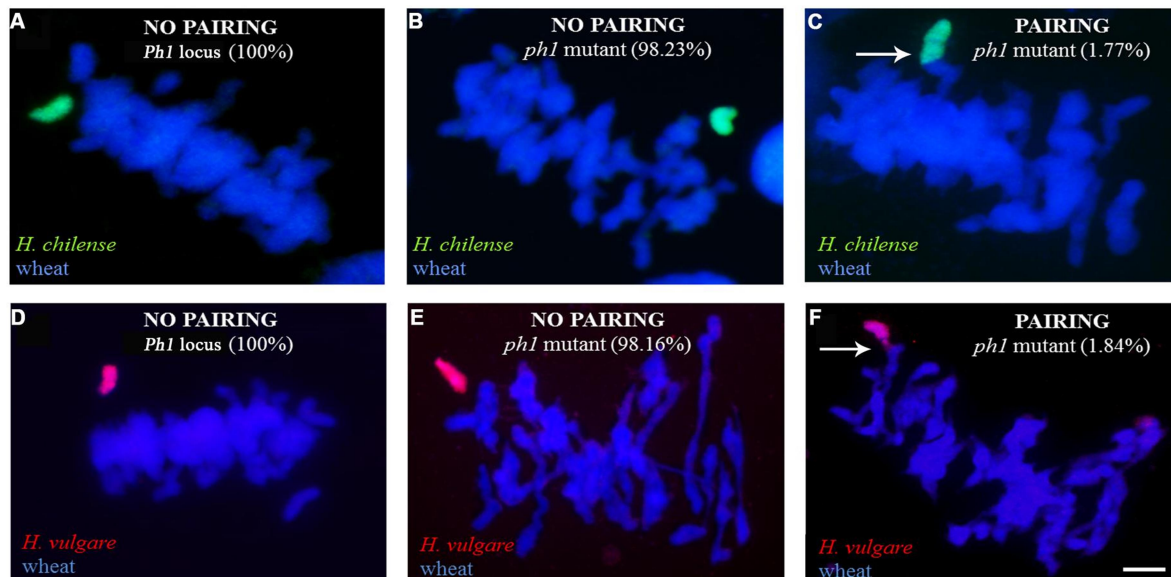


Figure 30 : Appariement des chromosomes à la métaphase I en fonction du statut allélique de *Ph1* (d'après [125])

Lorsque *Ph1* est présent, les chromosomes d'*Hordeum* sp. [(A) 7H<sup>ch</sup>, en vert et (D) 4H<sup>v</sup>, en rouge] restent non appariés. Dans le cas d'un fond génétique mutant *ph1b*, les chromosomes *Hordeum* sp. [(B) 5H<sup>ch</sup>, en vert et (E) 7H<sup>v</sup>, en rouge] restent à l'état d'univalent dans la plupart des cellules. Dans le cas de l'absence de *Ph1*, des appariements illégitimes ont lieu entre les chromosomes *Hordeum* sp. [(C) 5H<sup>ch</sup>, en vert et (F) 7H<sup>v</sup>, en rouge], et les chromosomes homéologues du blé. Echelle : 10 µm.

La densité en gènes est supérieure dans les régions distales avec notamment une proportion plus grande de gènes non synténiques [122]. De plus, l'étude de l'expression des gènes le long du chromosome 3B à l'aide de données de RNA-seq révèle également un partitionnement fonctionnel. Ainsi, l'expression des gènes constitutifs dans la région R2 est observée alors qu'il y a une expression des gènes spécifiques des organes dans les régions R1 et R3 [124].

En ce qui concerne les éléments transposables, le partitionnement est double. La région R2 est composée à 88% d'éléments transposables et les régions R1 et R2 à seulement 70% ce qui constitue un premier partitionnement quantitatif. Par ailleurs, la composition en types d'éléments répétés varie : les DNA-transposons sont présents majoritairement dans les régions R1 et R3 [123] alors que les rétro-transposons prédominent dans la région R2.

## 5. Contrôle de la recombinaison homéologue

En dépit du fait qu'il possède sept jeux de chromosomes très similaires (homéologues), le blé tendre se comporte comme une espèce diploïde lors de la méiose ce qui implique un niveau de régulation supplémentaire pour éviter leur appariement. Dès les années 50 [126], il a été montré que l'appariement correct des chromosomes homologues est contrôlé par un locus majeur appelé *Ph1* (pour Pairing homoeologous 1) localisé sur le bras long du chromosome 5B et que son effet est particulièrement visible dans des hybrides blé/seigle [127] (Figure 30). L'absence de *Ph1* chez les espèces diploïdes apparentées suggère que ce locus est apparu suite aux événements successifs de polyploïdisation [128]. Le clonage positionnel récent de ce locus [129] montre une structure complexe consistant en un segment d'hétérochromatine provenant de la partie subtélomérique du chromosome 3B insérée dans un bloc de cinq gènes *CDK2-L*. L'analyse cytogénétique d'hybrides blé/seigle dans un contexte sauvage ou mutant pour *Ph1* montre que ce locus stabilise la méiose polyploïde en favorisant l'appariement homologue et en empêchant simultanément les sites *MLH1* entre chromosomes homéologues d'être résolus en crossing-overs [130].



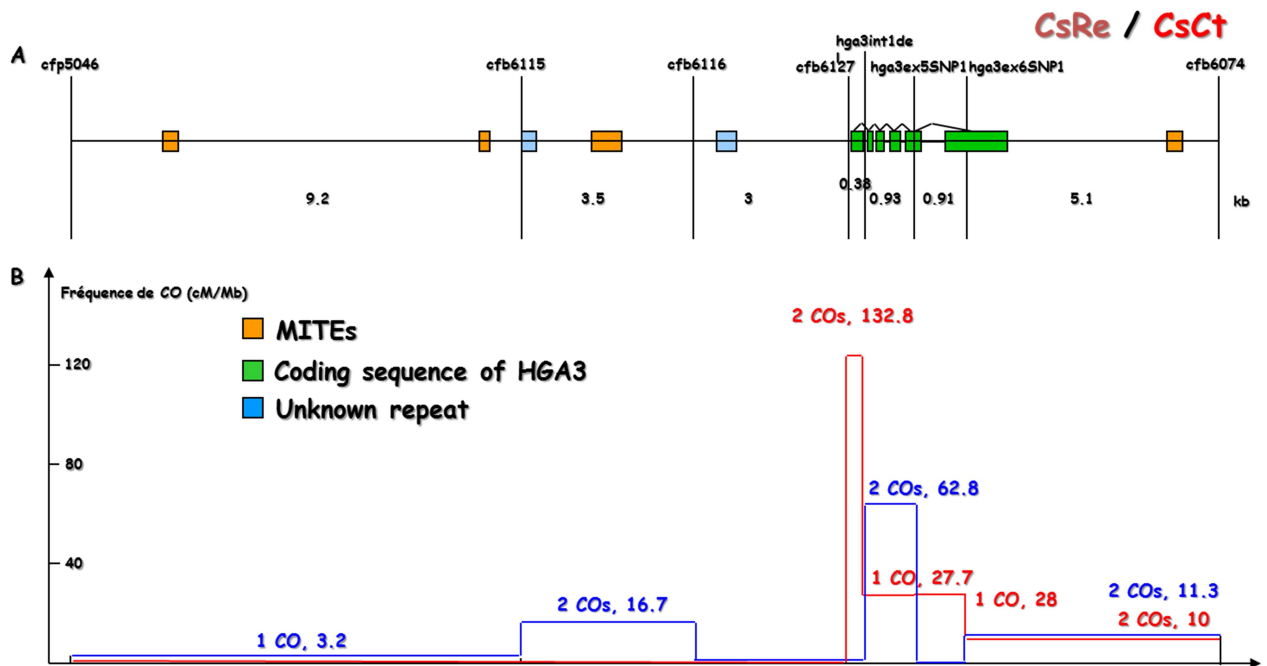


Figure 31 : Distribution de la recombinaison dans la région du gène *HGA3* (d'après [76])

Distribution des crossing-overs dans une fenêtre de 23kb qui entoure la région du gène *HGA3* entre les marqueurs cfp5046\_aspcr40 et cfb6074. (a) Les exons du gène *HGA3* sont représentés avec des boîtes vertes. Les boîtes orange et bleue représentent respectivement les éléments transposables de type MITE et les répétitions inconnues. La taille respective de chaque intervalle est indiquée (en kb). (b) La variation du taux de crossing-overs pour la population F2-CsRe (en bleu) et F2-CsCt (en rouge). Le nombre de crossing-overs et la fréquence (en cM/Mb) sont indiqués pour chaque intervalle.

En plus de *Ph1*, Le groupe d'homéologie 3 joue un rôle dans la recombinaison et l'appariement des chromosomes chez le blé. En effet, les lignées nullisomiques pour le chromosome 3B (N3B) produisent lors de la méiose un nombre anormalement élevé d'univalents [131] et ont permis la formation de toutes autre lignées aneuploïdes, ce qui montre son importance pour la bonne ségrégation des chromosomes. Le locus le mieux connu porté par le groupe d'homéologie 3 est le locus *Ph2* localisé sur le bras cours du chromosome 3D (3DS ; [132–134]). Ce gène n'a pas encore été identifié mais il affecte la recombinaison homéologue seulement chez les blé allo-haploïdes (ABD), pentaploïdes (AABBD) et les hybrides haploïdes blé-seigle (ABDR) [132,133]. Cependant, bien que le retrait de *Ph2* n'affecte pas le taux de recombinaison chez le blé lui-même [135,136], des études supplémentaires suggèrent que *Ph2* peut jouer sur la recombinaison homologue ainsi que sur la progression de l'appariement [137]. Un autre gène qui préviendrait la recombinaison homéologue est aussi mentionné sur le bras court du chromosome 3A (3AS) [134,138]. Ce gène serait moins efficace que *Ph2* [138] mais ça position relative suggère qu'il pourrait être une copie homéologue de *Ph2*. Les expériences réalisées avec des hybrides blé-seigle [134] montrent que le retrait des bras 3AS ou 3DS engendre un niveau d'appariement homéologue au moins équivalent à celui des plantes où *Ph1* est absent.



## 6. Etude des points chauds de recombinaison

Une seule étude, a pour l'instant été réalisée à une échelle fine chez le blé tendre. Celle-ci a porté sur un contig de 3Mb environ situé sur le bras court du chromosome 3B analysé chez deux populations F2. L'analyse de 3 zones recombinogènes de ce contig montre que la taille cumulée des gènes est positivement corrélée avec le taux de recombinaison [76]. L'étude d'un point chaud particulier autour du gène *HGA3* (Figure 31) suggère que les CO interviennent plus fréquemment dans les parties promotrices des gènes. Ces résultats montrent que chez le blé, comme chez *Arabidopsis* [77] et chez beaucoup d'autres espèces [73,75,82,139–141], la recombinaison a lieu à proximité des gènes ce qui suggère des paramètres communs pour la formation des crossing-overs.

Cette étude a également permis de mettre en évidence une conservation des points chauds entre les deux populations qui partagent un même parent ce qui suggère que les points chauds de recombinaison chez les plantes évoluent moins vite que chez les espèces possédant le système PRDM9 [48].

## 7. Analyse du déséquilibre de liaison chez le blé

Chez le blé tendre des études portant sur le déséquilibre de liaison (DL) montrent que celui-ci diminue dans les parties distales des chromosomes, où la recombinaison se situe majoritairement, alors qu'il reste élevé dans les zones ne recombinant que très peu [121,142]. L'étude du déséquilibre de liaison sur le même contig que précédemment à une échelle fine a permis de montrer une chute importante du DL dans le corps du gène *HGA3* et la mise en évidence d'un événement de conversion génique soulignant que ces phénomènes existent également chez le blé [87].



# I. Objectifs de la thèse

---

Chez le blé tendre, la méiose et la recombinaison restent mal connues alors que ces processus revêtent une importance primordiale pour l'amélioration de cette espèce agronomique. Ceci est dû en grande partie à la complexité et à la taille de son génome qui ont longtemps retardé le développement d'outils moléculaires adéquats. La création en 2005 du consortium international pour le séquençage du génome de blé (IWGSC, International Wheat Genome Sequencing Consortium ; <http://www.wheatgenome.org/>) a permis de lever les obstacles et d'accéder aux premières données de séquence partielles [143] et très récemment complètes avec à la fin 2016, la séquence du génome du blé tendre assemblée en pseudo-molécules grâce aux nouveaux algorithmes développés par la société NRGene®.

Les objectifs de la thèse étaient donc d'exploiter ces nouveaux outils pour étudier si **des caractéristiques de la séquence pouvaient expliquer la localisation des points chauds de recombinaison ainsi que la répartition du déséquilibre de liaison le long du chromosome 3B, chromosome modèle du blé tendre.** Ce choix de chromosome est justifié par : (1) sa taille qui permet de l'isoler par cytométrie de flux [144] ; (2) les nombreux gènes d'intérêt agronomique qu'il porte [145] ; (3) sa bonne relation de colinéarité avec le chromosome 1 du riz [146]. Ce travail a été organisé en trois parties :

## *Partie I :*

Une étude préliminaire réalisée au laboratoire sur des lignées de délétions du chromosome 3B a montré que la suppression totale ou partielle du bras long de ce chromosome impactait significativement le nombre de chiasmas et suggérait l'existence d'au moins deux locus différents affectant ce caractère (données internes au laboratoire non publiées). Avec l'accès à la séquence du chromosome 3B et la génération de données de RNA-seq dérivées d'anthères en développement à différents stades, plusieurs gènes candidats ont été postulés.



L'hypothèse de ce déficit en chiasmas est une mauvaise mise en place de l'axe synaptonémal. Avec l'appui de la DARESE dans le cadre de l'EIRA, j'ai pu réaliser un stage de 3 mois au sein du James Hutton Institute en Ecosse encadrée par Isabelle COLAS pour mener une étude cytogénétique sur les lignées de délétion les plus affectées. L'objectif était : (1) de transférer les techniques de cytogénétique de l'orge vers le blé ; (2) d'observer la mise en place de la synapse chez le blé tendre (Chinese Spring) ; (3) d'observer le comportement méiotique des lignées de délétion.

## **Partie II :**

Grâce aux nouvelles technologies de séquençage, de nombreux marqueurs génétiques de type SNP (plusieurs centaines de milliers) ont été détectés sur les 21 chromosomes du blé tendre. Plus de 420 000 marqueurs ont fait l'objet d'une validation *in vivo* à l'aide d'une puce Axiom® 420k développée dans le cadre du Projet Investissement d'Avenir BreedWheat (coordinateur du projet J Le Gouis, équipe BIANCA de l'Unité d'accueil ; coordinateur du développement de la puce E Paux, équipe SEVEN de l'Unité d'accueil). Cette puce a été utilisée pour génotyper une partie de ma population d'intérêt (Chinese Spring X Renan ; lignées SSD en génération F6) qui comporte plus de 1200 individus. Ainsi, j'ai réalisé la cartographie génétique génome entier du blé tendre ce qui m'a permis d'étudier les profils de recombinaison au niveau du génome entier et de mener une analyse comparative entre les chromosomes homéologues. Outre cet intérêt, cela m'a fourni un jeu supplémentaire de données pour l'analyse fine de la recombinaison aux locus sélectionnés.





### *Partie III :*

Enfin, chez le blé tendre un seul point chaud de recombinaison localisé dans la partie 5' du gène *Hga3* a été étudié [76]. Cette étude montre que la taille physique cumulée des gènes dans un contig a un rôle majeur dans la détermination de la fréquence des événements de recombinaison dans la région considérée et qu'il existe un lien entre la présence des gènes et la recombinaison chez le blé tendre. Par ailleurs, cette étude a pu, grâce au déséquilibre de liaison montrer un événement de conversion génique [76]. Cependant, les conclusions établies à partir d'un seul événement ne peuvent être généralisées et il est essentiel d'analyser plus de points chauds pour progresser dans ce domaine.

Dans ce but, j'ai d'abord, réalisé le génotypage de la population complète (1270 individus) Chinese Spring X Renan afin de détecter les lignées recombinantes aux locus les plus prometteurs (taux de recombinaison élevé). Pour continuer, j'ai réalisé une densification en marqueurs génétiques afin de pouvoir localiser finement les événements de recombinaison dans des fenêtres de quelques kilobases pour étudier les caractéristiques des séquences qui entourent les événements de recombinaison en termes de pourcentage de GC, contenu en gènes, motifs particuliers, etc...

Dans un second temps, grâce aux données de génotypage de la collection de blé tendre réalisée dans le cadre du projet BreedWheat, deux sous-populations ont été constituées pour étudier la recombinaison ancestrale dans deux pools génétiques très contrastés. Cette étude du déséquilibre de liaison est complémentaire à l'étude des points chauds de recombinaison et va permettre de confronter les données de recombinaison actuelle aux observations de la recombinaison ancestrale permises par l'étude du déséquilibre de liaison.

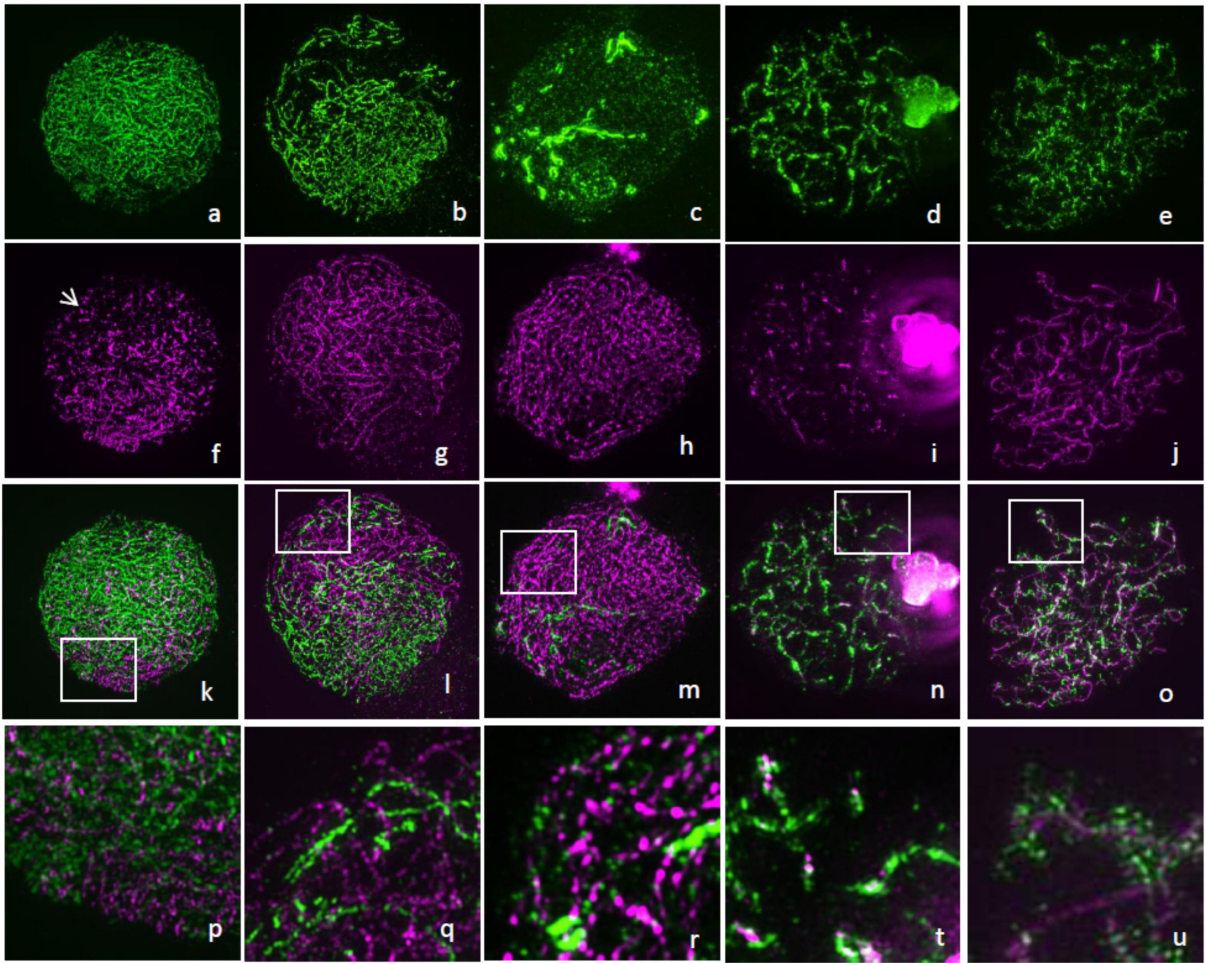


Figure 32 : Immunolocalisation d'ASY1 (en vert) et de ZYP1 (en magenta) durant la prophase I de méiose chez des méiocytes de blé de Chinese Spring.

Initiation de l'appariement au leptotène (a,f,k et p), progression de l'appariement en milieu de zygotène (b,g,l etq) and zygotène tardif (c,h,m et r), appariement complet aux pachytène (d,i,n et t) et stade 'tinsel' chromosome au diplotène (e,j,o et u).

## II. Partie I – Cytogénétique :

---

### 1. Transfert des protocoles de cytogénétique de l'orge vers le blé

Le James Hutton Institute en Écosse, a développé des protocoles expérimentaux chez l'orge qui permettent de réaliser de l'immunofluorescence tout en conservant la visualisation 3D des noyaux en méiose. Dans un premier temps, il était nécessaire de transférer les techniques de l'orge vers le blé. Pour cela, nous avons réalisé une série d'expérimentations sur méiocytes de blé en développement à l'aide des anticorps d'orge (*ASYI* et *ZIP1*) afin d'observer le déroulement de la méiose chez le blé (Chinese Spring). Les résultats obtenus font l'objet d'un article en cours de finalisation dans lequel la mise en place de la synapse chez l'orge et le blé est comparée. Pour des raisons de confidentialité de certains résultats, il n'est pas possible de présenter l'article dans son intégralité. Néanmoins, le protocole est détaillé dans l'article présenté dans la partie II qui concerne la recherche de gènes impliqués dans le comportement méiotique sur le chromosome 3B. Seul le déroulement du synapsis chez le blé est présenté ici.

En utilisant les anticorps TaASYI contre les protéines de blé et HvZYP1 contre les protéines d'orge, nous avons été capables de suivre le synapsis dans les méiocytes de blé (Figure 32) en utilisant la technologie 3D-SIM. Les télomères du blé forment d'abord un bouquet d'un côté du noyau où un intense remodelage se produit pour faciliter l'appariement et la recombinaison [147–149]. Aux stades leptotène/zygotène (Figure 32, a,f,k et p), TaASY1 se localise tout le long du chromosome avec une apparence granulaire similaire à l'orge. Au début du synapsis, HvZYP1 se localise au niveau du bouquet télomérique (Figure 32 f,k et p) attirant les chromosomes homologues ensemble jusqu'à leurs extrémités.



La présence de foci additionnels à l'intérieur du noyau suggère que les sites d'appariement sont aussi initiés tout le long des chromosomes (Figure 32 f, flèche) et confirme le processus dit « en bouton pression » (initiation en plusieurs sites le long du chromosome et entre les sites des espaces vacants) déjà décrit chez le blé [149]. Au milieu du Zygotène, la moitié du noyau est apparié (Figure 32 b,g,l et q), les cellules sont toujours polarisées et TaASY1 (Figure 32 b) n'est pas visible dans les zones de fort signal d'HvZYP1 (Figure 32g). Cependant, il n'est pas possible de voir la structure tripartite du complexe synaptonémal (Figure 32 q) ce qui suggère que la chromatine est très condensée et n'est pas accessible pour les anticorps. Au zygotène tardif, la plupart du noyau est maintenant appariée (Figure 32 c,h,m et r) et les points révélés par TaASY1 (Figure 32c et r) sont visibles le long du signal de la structure linéaire HvZYP1 ce qui suggère un appariement complet et un changement de l'état de condensation. Ce changement pourrait permettre de nouveau aux anticorps d'accéder à la protéine TaASY1. Néanmoins, il n'est toujours pas possible de voir la structure tripartite du complexe synaptonémal (Figure 32 m et r) ce qui suggère que ce dernier n'est pas finalisé, que la chromatine est trop condensée et/ou que les chromosomes sont très enroulés. L'appariement est complet à la fin du pachytène avec un signal épais de HvZIP1 tout le long des chromosomes (Figure 32 n). La protéine TaASY1 enrobe ceux-ci qui restent maintenus ensemble par HvZYP1. Elle tourne, s'élargie et émerge du cœur de l'axe des éléments latéraux (Figure 32 d et n). Il n'est toujours pas possible de voir une structure tripartite du complexe synaptonémal excepté dans quelques régions du noyau (Figure 32t) ce qui suggère une fois encore un niveau plus élevé de condensation. Finalement, au stade diplotène (Figure 32 e,j,o et u), lorsque HvZYP1 est sous forme linéaire, TaASY1 s'organise en branches qui se situent à l'extérieur le long du cœur du chromosome (Figure 32 u).



## **2. Recherche de gènes impliqués dans le comportement méiotique sur le chromosome 3B**

Le chromosome 3B qui sert de modèle pour les études de la distribution de la recombinaison (voir § suivants) porte également des gènes impliqués dans le déroulement de la méiose. En effet, la suppression complète de ce chromosome (plante nulli-3B ; N3B) entraîne la production de lignées chez lesquelles certains chromosomes sont absents. C'est cette particularité qui a permis à Sears (1944) de développer les stocks complets de lignées aneuploïdes (nullisomiques, nulli-tétrasonomiques, ditélosomiques) dans la variété Chinese Spring qui est actuellement la variété de référence internationale séquencée.

L'objectif de cet article en cours de finalisation était donc d'exploiter les techniques acquises dans le cadre de mon séjour en Ecosse pour caractériser plus précisément où sont localisés les gènes impliqués dans ce comportement et d'essayer d'identifier des candidats potentiels par analyse bioinformatique de données d'expression des gènes méiotiques.





# Location and identification of genes impacting chiasmata number on chromosome 3B of bread wheat

---

## Authors

Darrier Benoit<sup>1,\*</sup>, Colas Isabelle<sup>2,\*</sup>, Gay Georges<sup>1,\*</sup>, Rimbert H el ene<sup>1</sup>, Choulet Fr ed eric<sup>1</sup>, Pingault Lise<sup>1</sup>, Perretant Marie-Reine<sup>1</sup>, Paux Etienne<sup>1</sup>, Bernard Michel<sup>1</sup>, Jenczewski Eric<sup>3</sup>, Feuillet Catherine<sup>1,4</sup>, Sourdille Pierre<sup>1</sup>

## Affiliations

<sup>1</sup>GDEC, INRA, UBP, 63039, Clermont-Ferrand, France

<sup>2</sup>The James Hutton Institute, Dundee DD2 5DA, Scotland UK

<sup>3</sup>IJPB, INRA, 78000 Versailles, France

<sup>4</sup>Present address: Bayer Crop Science, Morrisville, NC 27560, USA

\*: the three first authors contributed equally to this work.

Running Title: Genes impacting chiasma in Wheat.

Key words: Wheat; Chiasmata number; Physical mapping; Meiosis; Candidate genes; Deletion bin

Corresponding author: Pierre SOURDILLE

*Genetic, Diversity and Eco-physiology of Cereals, INRA, Site de Crou el, 5 Chemin de Beaulieu, 63039, Clermont-Ferrand, France*

Phone: 33 4 73 62 44 37

Fax: 33 4 73 62 44 53

E-mail: [pierre.sourdille@clermont.inra.fr](mailto:pierre.sourdille@clermont.inra.fr)



## 1. Introduction

Meiosis is a specialized cell division which acts to halve the chromosome complement, or ploidy, during the production of gametes for sexual reproduction in eukaryotes. To ensure that each gamete has a full complement of the genetic material, homologous chromosomes must pair and then separate in a coordinated manner during meiosis. With a few exceptions (*Drosophila*, *C. elegans*), this is mediated by recombination that reshuffles genetic material between chromosomes through the formation of genetic crossovers (COs; for a review see [16,150]). Moreover, homologous pairing, synapsis and recombination are interdependent [36,151]. Homologous chromosomes need first to search for their partner. Telomeres cluster during early meiosis to bring homologous chromosomes together and initiate synapsis [148,152]. When homologs are close enough this allows the formation of the synaptonemal complex (SC). During zygotene the two homologues progressively synapse along their entire length and the process is completed at pachytene [36,151]. In the meantime, pairing enables the initiation of recombination and when the SC is disassembled at diplotene, homologous chromosomes remain together by chiasmata (the cytogenetic manifestation of the crossovers COs) to ensure their alignment on the metaphase I plate. At metaphase I the bivalents align at the equatorial plate and each of the homologous chromosomes separates at anaphase I. A second round of cell division then follows, resulting in sister chromatid separation and the formation of haploid cells [153]. At least one CO is mandatory per homolog pair (obligate CO; [150]) and the absence of COs results in the random segregation of homologs at anaphase I and in the formation of aneuploid gametes at the end of meiosis.



Polyplod species tolerate quite well aneuploidy probably because of the redundancy of the genetic information carried by the homoeologous chromosomes. Among others, this is the case for hexaploid bread wheat (*Triticum aestivum* L.; AABBDD;  $2n=6X=42$ ) derived from two successive interspecific crosses [154–156] involving the three diploid species *T. monococcum ssp urartu* (AA), a yet unknown species related to the *Sitopsis* section (BB; the closest one being *Aegilops speltoides*; S genome) and *Aegilops tauschii* (DD). The complete set of aneuploid stocks for which each of the 21 chromosomes (or chromosome arms; [157,158]) is missing was produced [131] and further used for gene identification [159] as well as for molecular marker assignment [160,161]. Interestingly, the line missing chromosome 3B (nullisomic 3B) shows an aberrant meiotic behaviour (ref sears 1944) with numerous univalents and is thus giving a lot of aneuploid descents (monosomics or trisomics). This suggests that this chromosome is carrying essential genes involved in meiosis and that the 3B-homoeologous copy of these genes cannot be compensated by the other two chromosome-3A and -3D copies.

Chromosome 3B from bread wheat is the most studied since now more than 10 years. Because of its physical size, this was the first to be isolated through chromosome sorting and for which a BAC library was constructed [162] . Consequently, this was also the first for which an anchored physical map was elaborated [145] as well as a completely assembled and annotated pseudomolecule [121]. This unique resource was further used to study the global structure of the chromosome [121] . Regarding recombination distribution, the analysis shows that crossover events occur in only 13% of the chromosome mainly located in the two distal regions of 68 Mb and 59 Mb on the short and long arms, respectively [121].



On the contrary, a large proximal region of 649 Mb (84% of the chromosome) spanning the centromere shows almost no recombination (0.05 cM/Mb) while this region contains about 67% of the genes from chromosome 3B preventing thus an efficient admixture of alleles in this region. Moreover, the pseudomolecule allowed the evaluation of the content and distribution of transposable elements [123], the gene content and duplication [122] and their expression [124]. This resource is thus of main importance to identify the genes that impact recombination located on this chromosome.

The homoeologous group 3 including chromosome 3B is often stated as playing a role in recombination in wheat. The most well-known locus is *Ph2*, a yet unidentified locus located on the short arm of chromosome 3D (3DS; [132–134]) that affects homoeologous recombination only in wheat allo-haploids (ABD), pentaploids (AABBDD) and wheat-rye hybrids (ABDR) [132,133]. However, despite the fact that removing *Ph2* does not seem to affect recombination rate in wheat itself [135,136], additional evidences suggest that *Ph2* can act on homologous recombination as well [163] through synaptic progression [137]. Another gene preventing homoeologous recombination is also mentioned on 3AS [134,138].

The gene on 3AS is less effective than *Ph2* [138] but its position relative to the one of *Ph2* suggests that it could be an homoeologous copy of *Ph2*. Evidence from hybrids with rye [134] indicates that deficiency for both 3AS and 3DS results in a level of homoeologous pairing almost as high as that expected for plants missing *Ph1* (for Pairing homoeologous 1, [126], a gene involved in homoeologous pairing and recently cloned [129]). It was later suggested that *Ae. speltoides* affects homoeologous chromosome pairing in wheat by suppressing the activity of *Ph1* [164]. One suppressor with a major effect was mapped on the long arm of *Ae. speltoides* chromosomes 3S [165] and was named *Su1-Ph1*. This gene increased homoeologous recombination in *T. aestivum* x *Ae. speltoides* hybrids by 8.4 chiasmata/cell.





Finally, another gene preventing homoeologous recombination and that does not interact with *Ph1* was located on the long arm of chromosome 3B [166–168] as well as probable homoeoalleles on the long arms of chromosomes 3A and 3D [132,134,138].

Accessing to genes controlling recombination is essential for breeders in order to exploit the wealth of diversity that exists in wheat genetic resources and the homoeologous group 3 carries several genes that would be of main interest. The genomic resources available for chromosome 3B can boost research in this direction and help cloning the genes involved in recombination and located on chromosome 3B as well as on chromosomes 3A and 3D when homoeologous copies exist [132,134,138,157,165]. The objectives of this paper were to study the variation of chiasmata number in the set of deletion lines from chromosome 3B [158] in order to locate more precisely on this chromosome the genes that are involved in this trait. Reduction in chiasmata number in some of the deletion lines suggests the presence of at least two genes involved in this trait and located on the long arm of chromosome 3B. In a second step, we used the first draft of the sequence from chromosome 3B [121] to map *in silico* genes that are specifically expressed during meiosis ([169,170]; RNAseq data) and genes that are known to be involved in meiosis in model species in order to identify putative candidate genes implicated in recombination. We thus identified 8 genes located in the two deletion bins concerned that when deleted, could contribute to the reduction of chiasmata number on chromosome 3B by impacting genome stability and cell-cycle. In addition, we present original preliminary data on the meiotic behaviour in the two promising lines 3BL7-0.63-1.00 and C-3BL2-0.22.



## 2. Material and Methods

### A. Plant material:

To estimate the location of genes located on chromosome 3B involved in pairing of chromosomes at meiosis, we used the reference line Chinese Spring as control and the aneuploid stock of chromosome 3B described earlier [157–159,171] consisting in the nullisomic 3B (N3B), the nullisomic 3B-tetrasomic 3A (N3BT3A) line, two ditelocentric lines (Dt3BL and Dt3BS) and 14 chromosome-3B deletion lines. Each deletion line is defined by the fraction length of the arm that remains present on the deleted chromosome (for example line 3BS4-0.55 conserves proximal 55% and lacks the distal 45% of the short arm of chromosome 3B). These deletion lines define 16 deletion bins: 3BL7-0.63-1.00, 3BL10-0.50-0.63, 3BL9-0.38-0.50, 3BL1-0.31-0.38, 3BL8-0.28-0.31, 3BL2-0.22-0.28, C-3BL2-0.22, C-3BS5-0.07, 3BS5-0.07-0.33, 3BS1-0.33-0.55, 3BS4-0.55-0.56, 3BS2-0.56-0.57, 3BS9-0.57-0.75, 3BS7-0.75-0.78, 3BS8-0.78-0.87 and 3BS3-0.87-1.00. Each deletion bin is defined by the fraction lengths of two successive deletion lines (for example the deletion bin 3BS4-0.55-0.56 is the chromosome fragment covering the region between 55% and 56% of the short arm of chromosome 3B and delimited by deletion lines 3BS4-0.55 and 3BS2-0.56). Plants were grown in growth chambers until the first two leaves emerged. They were then potted and grown for two weeks in cool conditions (17°C day and 12°C night, 12 hours of day) and then for two months at 24°C day and 18°C night with 16 hours of day until the spikes arrive at the appropriate stage to observe chromosome pairing (metaphase).



## B. Pairing observation and statistical analysis:

Young spikes are collected and the correct (metaphase) meiotic stage is determined by evaluating one anther/flower through a rapid squash under carmine-acetic colouration. The two other anthers are fixed in a solution of alcohol/acetic acid (3/1) during 48h and are then dissected in aceto-carmine added with FeCl<sub>3</sub> using an Axioskop (Zeiss) dissecting microscope. Only pollen mother cells (PMCs) are conserved for further analyses. The slides are heated until separation of the chromosomes and are washed with acetic water (45%). They are then slightly pressed to spread out the chromosomes and they can be stored in the fridge for rapid observation or in alcohol at -20°C after removal of the slide.

Cytoplasmic proteins are removed with pepsin followed with a paraformaldehyde treatment and the slides are washed twice (2X SSC) and fixed in successive alcohol solutions (70%, 95% and 100%). Squashed anther preparations are then viewed using the Axioskop. To estimate chiasmata frequencies, 50 PMCs in metaphase I were scored for each plant and numbers of univalents, rod and ring bivalents and multivalents were recorded. To estimate the mean chiasmata frequencies, we assumed that univalents, rod bivalents, ring bivalents, trivalents, quadrivalents contribute respectively with zero, one, two, three and four chiasmata.

Statistical significance of the differences observed between the aneuploid lines and the normal Chinese Spring were done using a classical Student T test. Validity of the expected values was estimated with a  $\chi^2$  test.

## C. RNASeq data production and *in silico* gene mapping:

Total RNAs were extracted from four different stages of developing anthers (Latent/Leptotene; Zygotene/Pachytene; Diplotene/Diakinesis; Metaphase I) of the hexaploid wheat cv Chinese Spring and from 10 mg of anthers (30-50 anthers depending on the stage) using the Macherey-Nagel Nucleospin RNA-XS kit according to manufacturer instructions.



RNA-Seq non-oriented libraries were constructed in duplicates using the TruSeq kit (Illumina). The eight libraries were sequenced (GATC, Germany) on two lanes (four samples per lane) of HiSeq2000 (Illumina) with paired-end sequence (500 bp) in 2x100 bp which generated 40 to 50 million of pairs of reads per sample. For alignment of the reads, we used the sequences of the whole genome assembly NRGene and available through the International Wheat Genome Sequencing Consortium (<http://www.wheatgenome.org/>) and on 3B chromosome [121]. All the reads from the RNA-Seq libraries were mapped on the scaffolds representing. We used TopHat2 v2.0.8 (<http://tophat.cbc.umd.edu/>; [172]) with the default parameters except that we tolerated no mismatches or splice-mismatches. PCR duplicates that are generated during library construction were removed from the analysis using SAMTOOLS (rmdup option). Transcripts reconstruction and expression determination (FPKM; [173]) were analysed with Cufflinks v2.0.2 (<http://cufflinks.cbc.umd.edu/>). Using mapping of gene from Chromosome Survey Sequence [143] on whole genome assembly NRGene homeologous copie.

For candidate-gene identification, we selected the genes expressed during meiosis and a set of 106 genes (see supplementary Tab. S1) known to be involved in meiosis in Arabidopsis, rice, maize and *Aegilops*. We mapped all these genes onto the pseudomolecule of chromosome 3B [121] using their protein sequence determined through an exonerate (V2.2.0) analysis [174,175]. We confirmed their position using annotation already available and RNAseq data [121,124]. Only those showing the best hits for alignment (scores > 500, identity > 40%, similarity > 60%) and functional (no pseudo gene) annotated gene on pseudomolecule were considered as relevant. Only predictions with an FPKM value > 1 in at least one condition were considered for expression analysis. In addition, a comparative analysis of expression between homoeologous copies was conducted. Expression on homoeologous chromosomes 3A and 3D of genes identified on chromosome 3B was searched for using IWGSC survey protein database and blastp.





Homoeologous gene copies were selected according to the closest similarity with the copy from chromosome 3B using phylogenetic tree with 3B gene selected previously.

#### **D. Immunostaining:**

Staged anthers were collected in 1X-PBS containing 0.5% Triton® X100 in an embryo dish and subsequently fixed in freshly made 4% formaldehyde solution for 30min at room temperature. Fixed anthers were washed twice in 1X-PBS/0.5% Triton® X100 and tapped to release the meiocytes in the embryo dish. Meiocytes (30µL) were transferred onto a *Polysine® slide (Poly-L-Lysine coated slides)* and left to air-dry gently. Slides were first incubated for 1h at room temperature in 1X-PBS, 1% Triton® X100 for 30-45min to permeabilize the cells, then in a blocking solution consisting of 3% BSA in 1X-PBS, 0.1% TritonX100. Slides were incubated in the primary antibody solution in a wet chamber for 1h at room temperature followed by 1 or 2 days in the fridge (6-8°C). The primary antibody solution consisted of one or multiple antibodies previously described from the following list; anti-AtASY1 (rabbit, 1:5000) [176] and anti-HvZYP1 (rat, 1:500) (supplementary data Figure S4). After incubation, slides were let for an additional 1h at room temperature before washing for 15min in 1X-PBS and incubating for 2h at room temperature in the secondary antibody solution consisting of a mixture of anti-rabbit Alexa Fluor (488 or 568) and/or anti-rat Alexa Fluor (594 or 488) (Invitrogen) diluted in 1X-PBS (1:300). After a 15min wash in 1X-PBS, the slides were counterstained with DAPI (2µg/mL) for 10-15min, mounted in Vectashield (H-1000) and sealed with nail varnish.

#### **E. Fluorescence DNA *in situ* hybridization:**

The protocol is adapted from [177] and [152]. Briefly, anthers were fixed in Ethanol:Acetic Acid (3:1) for at least 24h and then transferred into 70% ethanol for storage. Anthers were spread on polysine slides according to Higgins et al (2012) and slides were air-dried.



The 45S probe was prepared from the 45S rDNA clone from wheat (pTa71; [178]) and labelled with biotin according to [177] and [152]. Hybridization mix, composed of 50% deionised formamide, 20% dextran sulfate, 1X PIPES/EDTA buffer (100:10), 0.3M NaCl, 500ng of salmon sperm blocking DNA, and 50ng of each probe, was denatured for 8min at 100°C and immediately placed on ice to cool down for 5min. Slides were successively dehydrated in 30, 50, 70 and 100% ethanol and 10uL of the probe mixture was applied onto the tissue, covered with a plastic coverslip and denatured on a hotplate for 5min at 75°C. Slides were moved into a wet chamber and incubated at 37°C overnight. The next day, slides were washed at 42°C in 20% formamide, 0.1X SSC for 10min, then in 2X SSC for 10min. This was followed by a wash at room temperature in 2X SSC for 10min, and in 4X SSC, 0.2% Tween 20 for 10min. A blocking solution, consisting of 5% BSA in 4X SSC, 0.2% Tween 20 was applied for 5min in a humidity chamber at room temperature. Biotin-labelled probes were detected with extravidin-cy3. Slides were counter stained in 1µg/mL Hoescht 33342.

#### **F. Microscopy and Imaging:**

3D-SIM acquisition of individual meiocytes was sequentially obtained using an OMX Blaze structured illumination microscope (GE Applied Precision) for three channels (405, 488 and 568 nm); SIM image reconstruction and processing was performed with the commercial software Softworx (GE Applied Precision) [179]. 3D Confocal stack images were acquired with LSM-Zeiss 710 for the three channels sequentially with an averaging of at least four, and de-convolved with the Tikhonov-Miller algorithm from the DeconvolutionLab package for Fiji/ImageJ. Projections of 3D pictures and light brightness/contrast adjustment were performed with the public domain program Fiji and/or Imaris (Bitplane).

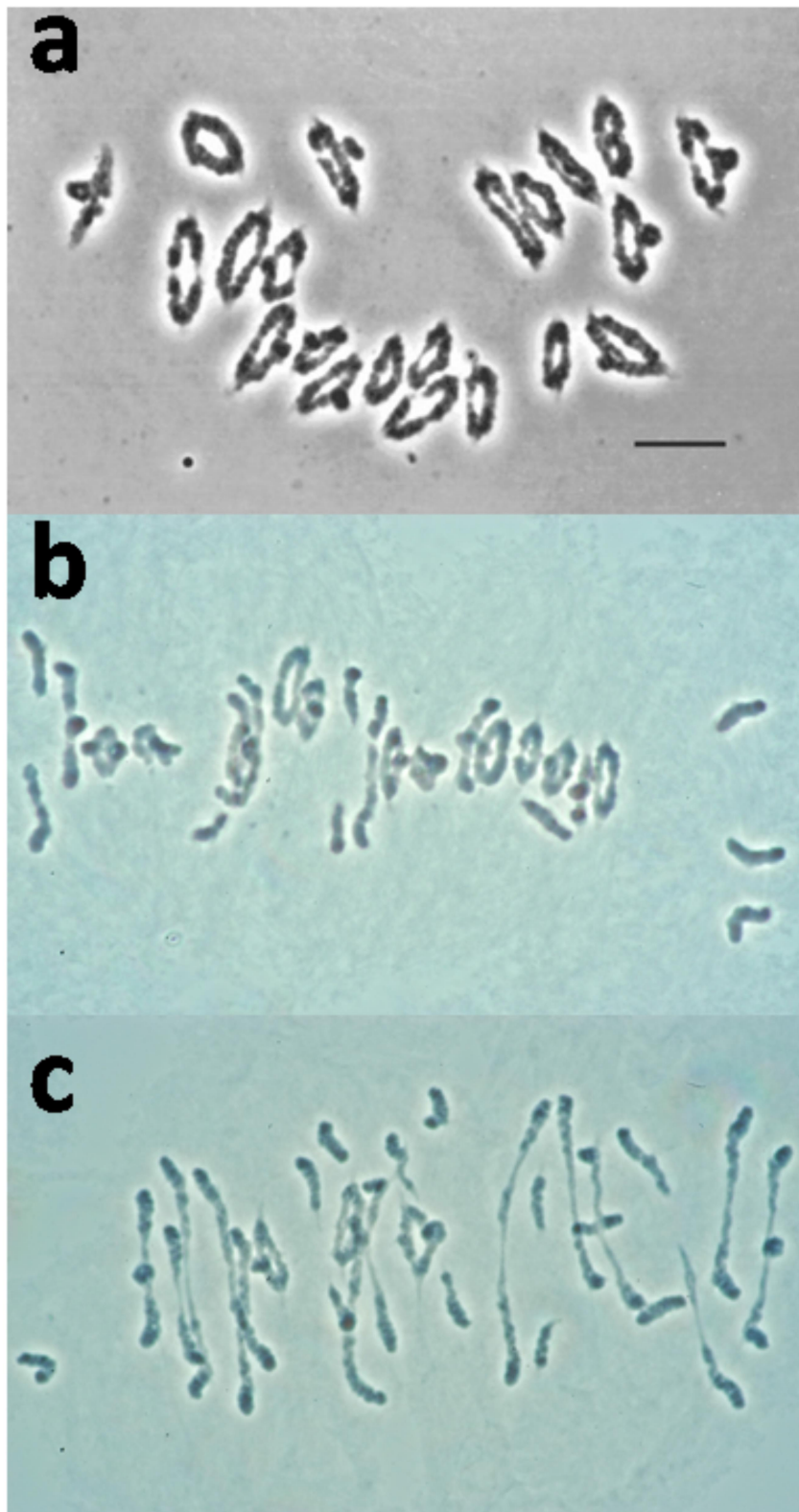


Fig. 1: meiotic behaviour of pollen mother cells of normal (a), Nulli 3B (b) and ditelosomic 3BS (c) Chinese Spring lines.

### 3. Results

#### A. The long arm of chromosome 3B is important for chromosome pairing.

Chromosome pairing was observed on a set of 50 meiotic (metaphase) cells and numbers of bivalents (rods and rings), univalents and chiasmata were counted. The wild type euploid CS ( $2n = 6x = 42$  chromosomes; 14 A, 14 B, 14 D; Fig. 1a) exhibits a mean of 21 bivalents ( $1.02 \pm 0.31$  rods and  $19.98 \pm 0.31$  rings) and of  $40.98 \pm 0.31$  chiasmata (Tab. 1). No univalent was observed which confirmed that as expected, euploid CS exhibits a normal meiotic pairing behavior. Similarly, the aneuploid lines missing either chromosome 3A (N3AT3B and N3AT3D) or chromosome 3D (N3DT3A and N3DT3B) showed pairing which was not statistically different from CS. On the contrary, the aneuploid line missing chromosome 3B (N3B; Fig. 1b) had more univalents and a significantly reduced number of chiasmata ( $35.22 \pm 0.55$ ), mainly because of an increased number of rod bivalents ( $4.34 \pm 0.48$ ). The aneuploid lines for which chromosome 3B is replaced either by chromosome 3A or 3D (N3BT3A and N3BT3D) revealed similar numbers of bivalents ( $19.46 \pm 0.32$  and  $19.70 \pm 0.25$  respectively) and intermediate numbers of chiasmata ( $36.74 \pm 0.54$  and  $37.32 \pm 0.52$ ) which indicates that the absence of chromosome 3B is partially compensated by supernumerary chromosomes 3A or 3D. Similar analyses were made on double ditelosomic lines ( $40''+2t$ ) missing either the short (named Dt3BL because only long arm (L) is present) or the long (Dt3BS) arms of chromosome 3B. For Dt3BL, we observed  $20.96 \pm 0.05$  bivalents ( $2.48 \pm 0.34$  rods and  $18.48 \pm 0.35$  rings) and  $39.44 \pm 0.36$  chiasmata (Tab. 1) which is statistically not different from the normal CS. This suggests that there is no gene involved in pairing on the short arm of chromosome 3B or that the gene(s) that could exist has(ve) only a limited effect on pairing. On the contrary, when Dt3BS (Fig. 1c) is analyzed, we observed means of only  $17.92 \pm 0.50$  bivalents ( $10.00 \pm 0.58$  rods and  $7.92 \pm 0.50$  rings) and  $25.84 \pm 0.81$  chiasmata.

Table 1: Chromosome pairing behaviour of 50 meiotic metaphases I cells for Chinese Spring and the aneuploidy line and deletion line of 3B chromosome.

Line	% arm	2n	Average of chiasmatas	Average of univalents	Average of bivalents	Average of rods	Average of rings	Average of multivalents
CS	1	42	40.98±0.31	0	21	1.02±0.31	19.98±0.31	0
N3B	0	40	35.22±0.55	0.44±0.23	19.78±0.11	4.34±0.48	15.44±0.50	0
N3AT3B	1	42	38.44±0.64	0.04±0.08	20.42±0.25	2.34±0.41	18.08±0.46	0.28±0.12
N3AT3D	1	42	38.82±0.57	0.28±0.19	20.22±0.26	2.90±0.53	17.32±0.54	0.32±0.13
N3BT3A	0	42	36.74±0.54	0.88±0.31	19.46±0.32	4.34±0.41	15.12±0.48	0.56±0.14
N3BT3D	0	42	37.32±0.52	0.52±0.24	19.70±0.25	4.08±0.52	15.62±0.60	0.52±0.14
N3DT3A	1	42	39.94±0.43	0.08±0.11	20.28±0.29	1.92±0.43	18.36±0.48	0.34±0.14
N3DT3B	1	42	39.82±0.41	0.08±0.11	20.24±0.26	1.98±0.38	18.26±0.43	0.36±0.13
Dt3BL	L	40+2t	39.44±0.36	0.08±0.11	20.96±0.05	2.48±0.34	18.48±0.35	0
Dt3BS	S	40+2t	25.84±0.81	6.16±0.99	17.92±0.50	10.00±0.58	7.92±0.50	0
F1CSx3BS	1	41+t	36.52±0.52	2.00±0.40	19.84±0.27	4.16±0.48	15.68±0.51	0.44±0.14
F1CSx3BL	1	41+t	39.30±0.30	0.16±0.15	20.92±0.08	2.54±0.30	18.38±0.29	0
3BS3	0.87	42	39.68±0.37	0	21	2.32±0.37	18.68±0.37	0
3BS8	0.78	42	39.92±0.45	0.12±0.13	20.94±0.07	1.96±0.43	18.98±0.43	0
3BS7	0.75	42	40.40±0.33	0.12±0.13	20.94±0.07	1.48±0.32	19.46±0.32	0
3BS9	0.57	44	40.00±0.52	0.44±0.26	21.78±0.13	3.56±0.47	18.22±0.48	0
3BS2	0.56	42	40.36±0.32	0	21	1.64±0.32	19.36±0.32	0
3BS4	0.55	42	39.78±0.42	0.04±0.08	20.98±0.04	2.18±0.41	18.80±0.41	0
3BS1	0.33	42	40.20±0.27	0.04±0.08	20.98±0.04	1.76±0.26	19.22±0.26	0
3BS5	0.07	42	39.64±0.41	0.08±0.16	20.96±0.08	2.28±0.37	18.68±0.38	0
3BL2	0.22	42	34.00±0.68	2.20±0.52	19.90±0.26	5.80±0.48	14.10±0.53	0
3BL8	0.28	42	37.54±0.51	0.36±0.21	20.82±0.11	4.10±0.46	16.72±0.47	0
3BL1	0.31	42	36.80±0.59	1.22±0.31	20.20±0.23	4.52±0.54	15.68±0.56	0.46±0.14
3BL9	0.38	41	33.10±0.65	2.18±0.41	19.38±0.23	5.70±0.50	13.68±0.55	0.02±0.04
3BL10	0.50	42	35.90±0.69	1.04±0.40	20.44±0.21	5.04±0.55	15.40±0.59	0.02±0.04
3BL7	0.63	41	33.46±0.71	2.12±0.34	19.44±0.17	5.42±0.58	14.02±0.63	0

Especially, the number of rod bivalents was largely increased (10 times) compared to normal CS. At the same time, the mean number of univalents was  $6.16 \pm 0.99$  while none was observed in euploid CS. These results suggest that at least one gene with a strong effect on pairing locates on the long arm of chromosome 3B. We then analyzed the meiosis of F1 hybrids derived from the cross between CS and either Dt3BL or Dt3BS. In the first case, two doses of 3BL for one of 3BS showed a mean number of chiasmata ( $39.30 \pm 0.30$ ) similar to what is observed with Dt3BL ( $39.44 \pm 0.36$ ). Absence of difference confirms that no gene with a significant effect is present on 3BS. On the contrary, one dose of 3BL for two doses of 3BS showed a mean frequency of chiasmata ( $36.52 \pm 0.52$ ) intermediate between those of Dt3BS ( $25.84 \pm 0.81$ ) and CS ( $40.98 \pm 0.31$ ) which suggests an additive effect of the gene(s) that is(are) present on this arm. From these results, we can conclude that at least one gene exists on the long arm of chromosome 3B involved in pairing and having a strong additive effect. We thus decided to locate this(ese) gene(s) more precisely by evaluating chromosome pairing using the deletion lines from CS.

### **B. The long arm of chromosome 3B bears at least two genes involved in pairing between chromosomes**

Eight deletion lines missing from 13% (3BS3) to 93% (3BS5) of the short arm of chromosome 3B were analyzed. In all these line, we observed almost no univalent ( $0$  to  $0.44 \pm 0.26$ /cell; Tab. 1; Fig. 2), an average number of bivalents ranging from  $20.94 \pm 0.07$  (3BS7, 3BS8) to  $21$  (3BS2, 3BS3) and a slight variation of number of chiasmata between the lines, from  $39.64 \pm 0.41$  (3BS5) to  $40.40 \pm 0.33$  (3BS7). However, these differences were not statistically significant compared to the wild type CS ( $21$ ). As we previously suggest from the Dt3BL pairing result, this confirmed that if genes involved in pairing exist on 3BS, they must have a small effect.



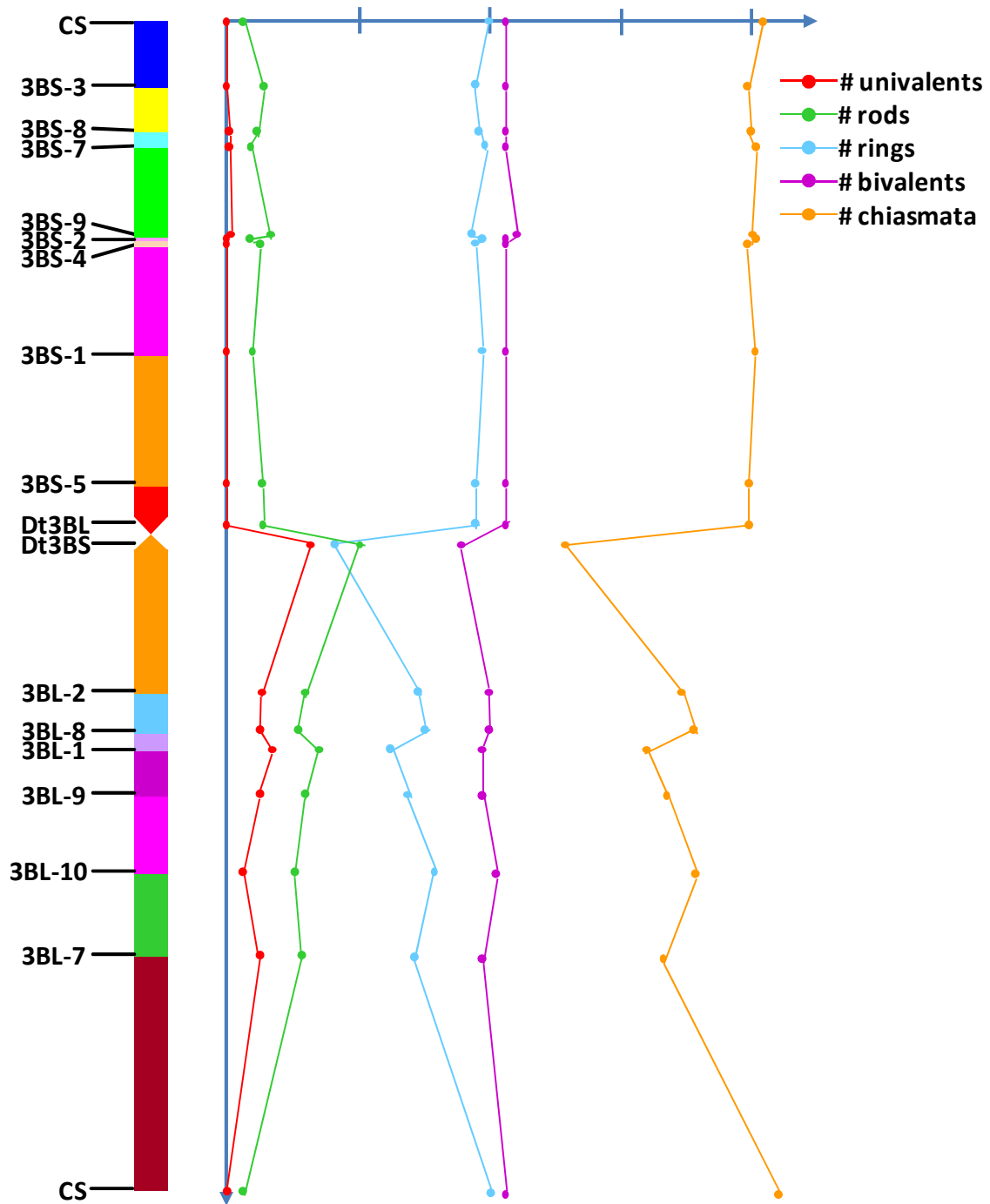


Fig. 2: Mean numbers of univalents, bivalents (rod and ring) and chiasmata for the various Chinese Spring lines: euploid (CS), ditelosomic (Dt3BS and Dt3BL) and deletion lines. Deletion bins are colour-coded on the left of the figure and are represented at the scale.

Similarly, six deletion lines missing from 37% (3BL7) to 78% (3BL2) of the long arm of chromosome 3B were analyzed. They had similar averages of bivalents ( $19.38 \pm 0.23$  –  $20.82 \pm 0.11$ ) as the N3B line ( $19.78 \pm 0.11$ ). Contrary to the 3BS deletion lines, high numbers of univalents were observed for all the 3BL deletion lines ranging from  $0.36 \pm 0.21/\text{cell}$  (3BL8) to  $2.20 \pm 0.52/\text{cell}$  (3BL2; Tab. 1; Fig. 2). However, this remained at least twice lower than what was observed for the Dt3BS ditelosomic line missing the entire long arm ( $6.16 \pm 0.99/\text{cell}$ ). The main difference between the deletion lines and the normal CS was due to larger numbers of rod bivalents ( $4.10$ - $5.80/\text{cell}$  compared to  $1.02/\text{cell}$  respectively). This leads to a significant decrease of the numbers of chiasmata for all the deletion lines which ranged from  $33.10 \pm 0.65$  (3BL9) to  $37.54 \pm 0.51$  (3BL8; Tab. 1; Fig. 2). Comparative analysis between the normal CS, deletion lines from the long arm and with the ditelosomic line Dt3BS indicated that the deletion lines had similar number of chiasmata but that the difference was significant with both the euploid CS line and the Dt3BS (Tab. 1) which revealed that at least two genes are involved in pairing on this arm. The first is located distally in deletion bin 3BL7-0.63-1.00 and the second one is located proximally in deletion bin C-3BL2-0.22.

From the results of the analysis of the chromosome 3B deletion lines, we can conclude that: (1) no gene with a significant effect is located on the short arm; (2) the long arm of chromosome 3B bears at least two genes involved in pairing between chromosomes; (3) at least one of these genes has a strong effect.

Table 3: Related Meiotic gene detected in C-3BL2 and 3BL7 using models species protein and exonerate alignment (score >500 identity > 40% similarity > 60%). Related: name of related Arabidopsis's gene. Gene Id: Id of gene in Chromosome Survey Sequence[143]. Del.bin: Deletion bin. Fpkm Value for the 4 stage of meiosis available.

Related	DelBin	GeneId	Leptotene Zygotene (Fpkm value)	Pachytene Diotene (Fpkm value)	Diakinesis (Fpkm value)	Metaphase I (Fpkm value)
MPK	C-3BL2-0.22	Traes_3B_AD0872038	11.5	12.3	12.3	13.7
DUO1	C-3BL2-0.22	Traes_3B_D5B791086	0.9	1.8	1.8	1.3
RTEL1	C-3BL2-0.22	Traes_3B_1B3493811	1.7	2.6	2.6	1.8
CAP	3BL7-0.63-1.00	Traes_3B_8FCD39B95	7.8	8.2	8.2	6.3
MUS81	3BL7-0.63-1.00	Traes_3B_17863F332	5.2	6.5	6.5	5.4
MLH1	3BL7-0.63-1.00	Traes_3B_B2D317888	6.0	7.2	7.2	6.6
SYN4	3BL7-0.63-1.00	Traes_3B_5E153DF63	6.3	7.0	7.0	5.9
ZIP4	3BL7-0.63-1.00	Traes_3B_62B5CC26B	69.0	117.0	117.0	119.5

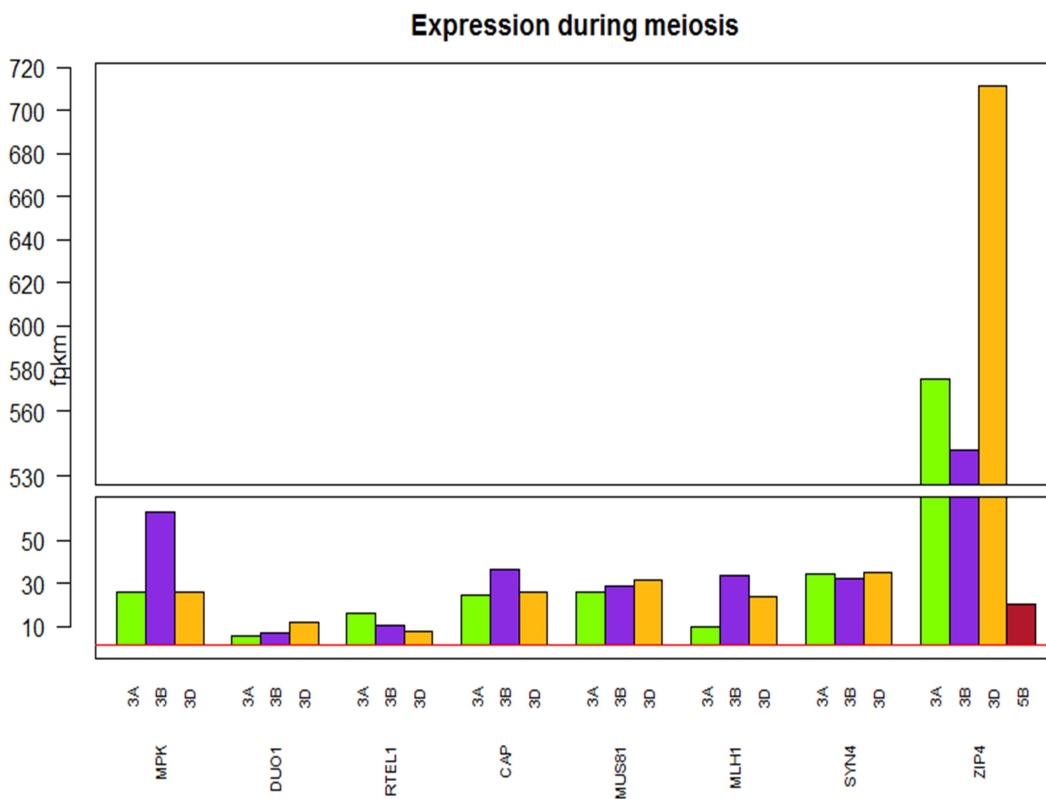


Fig. 3: Expression (fpkm) of homeologous gene detected on 3B chromosome long arms. Homeologous group are 3A (in green), 3B (in magenta), 3D (in orange) and 5B (in red)

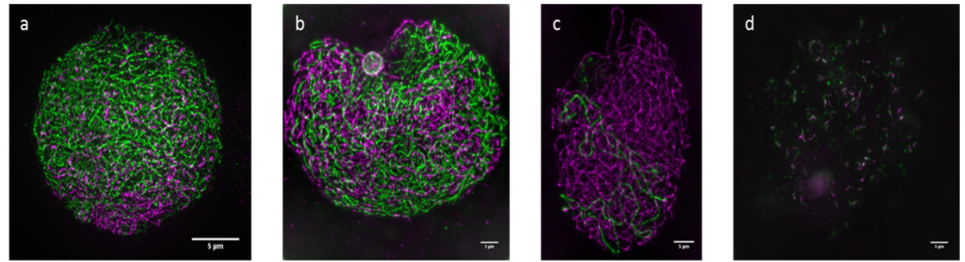
### C. *In silico* mapping of genes expressed during wheat meiosis or known as involved in meiosis in Arabidopsis

We used pseudomolecule of chromosome 3B [121] and RNASeq data from wheat developing anthers to identify genes expressed during meiosis and located in the various deletion bins. Among the 7264 genes from chromosome 3B, 5172 were placed in deletions bins (Tab. 2) and 3030 located on the long arm and 2142 on the short arm. RNA-seq data revealed 2611 genes (50%; 48% and 52% for short and long arm respectively) that were expressed during meiosis (FPKM > 1) and physically assigned to deletion bins. Number of expressed genes in the different deletion bins varied greatly and ranged from 7 (3BL8-0.28-0.32) to 672 (3BL7-0.63-1.00) probably because of a difference in bin sizes ranging from 157Mb (3BL7-0.63-1.00) to 7Mb (3BS4-0.55-0.57).

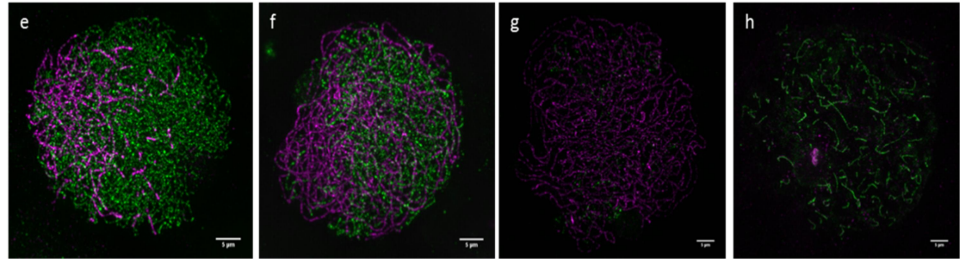
We also selected a set of 106 genes known to be involved in meiosis or playing a role in DNA repair (supplementary Tab. S1) to identify related gene on 3B. Eight related genes were mapped on the long arm (Tab. 3). Among the 8 genes, three are located in bin C-3BL2-0.22 (*RTEL1/DUO1/MPK* (Mitogen-activated protein kinase)) and five in bin 3BL7-0.63-1.00 (*MLH1/ZIP4/SYN4/MUS81/CAP*). According to the mapping, related *CAP* gene was found in only one copy in wheat chromosome 3B assembly.

Several of these genes could constitute good candidates and we analysed expression of their homoeologous copies using whole genome assemblies. From our RNA-Seq data, copies from chromosome 3B were the most expressed (Fig. 3) for *CAP* and *MLH1* in bin 3BL7-0.63-1.00 and *MPK* in bin C-3BL2-0.22. For *ZIP4*, chromosome-3B copy was less expressed and 3A and 3B were mainly expressed. Also, for *SYN4* and *MUS81* each of three copies were expressed at the same level. Herefore, from *in silico* analyses we can conclude that numerous genes expressed during meiosis locate in the two deletion bins of the long arm.

Chinese Spring



3BL-7-0.63-1.00



Dt3BS

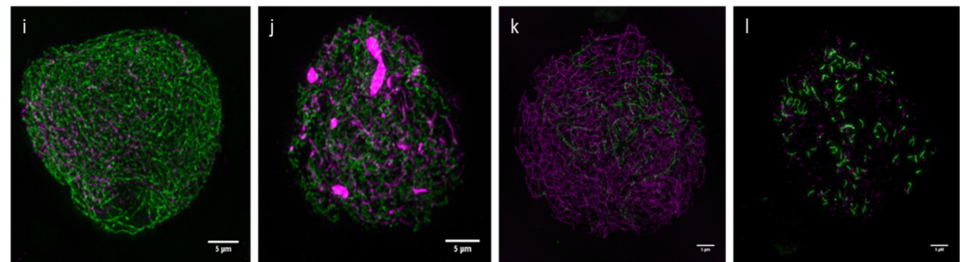


Fig 4 : progression of the synaptonemal complex during prophase in CS (a,b,c,d), 3BL-70.63-1.00 (e,f,g,h) et Dt3BS (i,j,k,l)

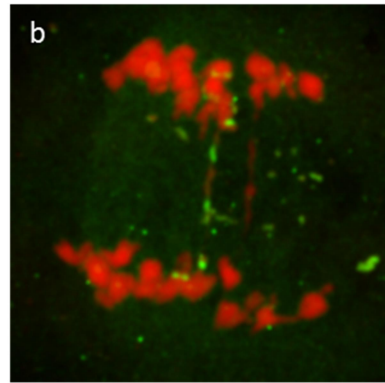
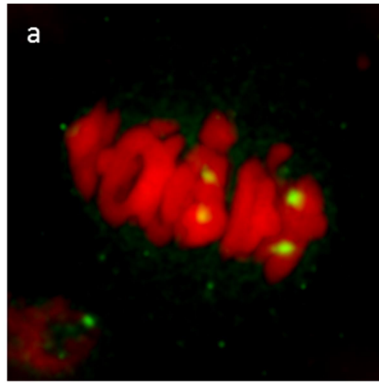
Immunolocalization of ASY1 (green) and ZYP1 (magenta) during meiotic prophase I. Initiation of synapsis at leptotene (a,e and i), synapsis elongation during zygotene (b,f and j), complete synapsis at pachytene (c,g and k) and late diplotene (d,h and l).

Some of these genes are known to be involved in meiosis (mismatch repair and genome stability) and can constitute good candidates that should be analysed in more details. For two of these genes already known to be involved in meiosis (CAP, MLH1, MUS81, SYN4), the chromosome-3B homeologous copie is the most expressed or expressed at similar level as the other homeologous copie.

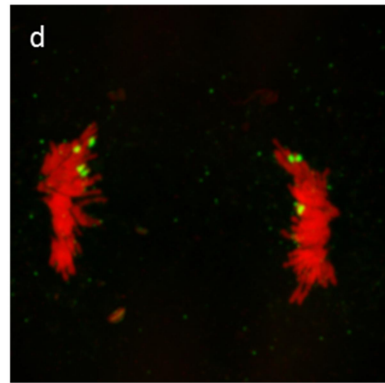
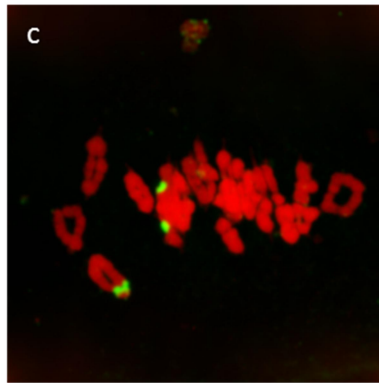
#### **D. Comparative analysis of meiotic behaviour in CS, Dt3BS and 3BL7-0.63-1.00.**

In order to refine our selection of candidate genes, we analysed the meiotic behaviour of lines Chinese Spring (CS; control), Dt3BS and 3BL7-0.63-1.00. Previous studies reported that Dt3BS was potentially asynaptic (no synapsis) and can present translocations [180]. Using antibodies raised against HvZYP1 and the axial element associated protein TaASY1 [176,181] combined to Structured Illumination Microscopy (SIM), we compared synapsis in CS, Dt3BS and 3BL7-0.63-1.00 (Fig 4). Axis formation, tracked by TaASY1 (green) appeared normal in all 3 lines as shown by the regular ASY1 signal (Fig. 4). At the onset of Zygotene, synaptonemal complex (visualized with HvZYP1 antibody - magenta) starts to polymerize at one side of the nucleus in CS (Fig. 4a), presumably at the telomere region, and brings the chromosomes in a zipper-like manner (Fig. 4b,c). At pachytene, the chromosomes are aligned along each other via the synaptonemal complex (Fig. 4d). The line 3BL7-0.63-1.00 displays normal synapsis from zygotene to pachytene (Fig. 4e-h) which was not significantly different from the wild type. In Dt3BS line, synapsis appears similar to CS at Leptotene/Zygotene (Fig. 4i) and at mid-zygotene (Fig. 4j). However, by late zygotene (Fig. 4k), Dt3BS exhibits ZYP1 polycomplex suggesting a defect in the maturation of the synaptonemal complex. At pachytene (Fig. 4l), although the chromosomes are fully aligned, ZYP1 is not fully polymerized revealing an abnormal synapsis. Contrary to what was hypothesized in previous study, Dt3BS is not asynaptic since we observed the formation of the synaptonemal complex (SC), although the SC is not normal.

Cs



3BL7



Dt3Bs

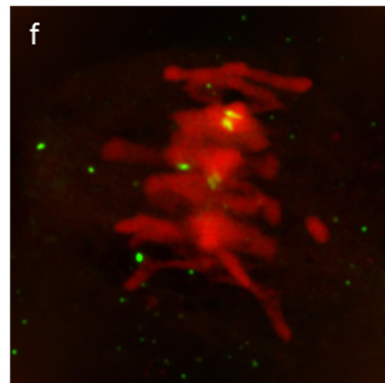
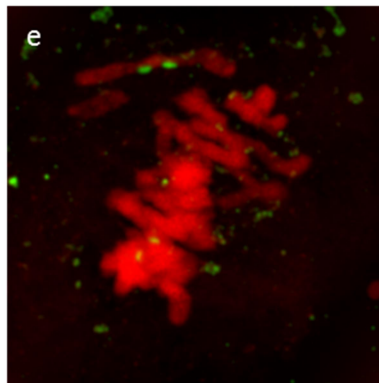


Fig 5: Spread of meiosis metaphase I/or telophase I of Cs (a,b), (c,d), (e,f). Chromosomes are labelled in red and the fluorescent DNA *in situ* hybridization of the rDNA probe 45S are labelled in green.

We also looked at possible non-homologous chromosome pairing using Fluorescent DNA *in situ* hybridization of the rDNA probe 45S (Fig. 5). We detected two large sites per homologous bivalents. In CS (Fig. 5a), we detected at least two ring bivalents carrying two 45S signal at the telomere region. In 3BL7-0.63-1.00 line (Fig. 5c), we also detected bivalents (ring or rod) carrying two signal for 45S and suggesting that in 3BL7-0.63-1.00 line, homologous chromosome pairing is not affected. Similarly, the Dt3BS line (Fig. 5e,f) seems to have normal homologous chromosome association. Chromosome segregation occurred normally in CS (Fig. 5b) and 3BL7-0.63-1.00 (Fig. 5d), despite the presence of occasional anaphase bridge in this latter. We were not able to find the corresponding Anaphase for Dt3BS as meiosis staging is generally difficult in this line. However, in Dt3BS, the chromosomes appeared fuzzier suggesting difficulties in chromatin compaction/condensation. In addition two nucleoli are observed at zygotene suggesting a delay in cell cycle. Altogether, our results suggest that the chromosome recognition is not affected in Dt3BS and 3BL7-0.63-1.00 and that the reduction of chiasmata (crossover) is likely due to an impair of the crossovers during meiosis.

## 4. Discussion

### A. Variation in chiasmata number in the aneuploid stock from chromosome 3B

Wheat is one amongst the allopolyploid species in which a substantial amount of work has been carried out to identify genes involved in the genetic control of pairing [129,176,182–185]. However, a lot of work remains to be done since many genes distributed throughout the genome are involved. Especially, chromosome 3B has been found to carry genes playing a role in pairing in many studies [157,166–168,186]. Fortunately, this chromosome is the one for which a lot of efforts have been done in order to obtain a high-quality sequence useful for wheat breeders [145,170] (<http://urgi.versailles.inra.fr/index.php/urgi/Projects/3BSeq>). In this study, we used the aneuploid stocks of wheat, namely the ditelosomic and deletion lines related to chromosome 3B to locate the genes involved in recombination and pairing on this chromosome.





In the euploid line Chinese Spring, we observed 21 bivalents corresponding to an estimated average of  $40.98 \pm 0.31$  chiasmata. According to the genetic map, barley (*Hordeum vulgare*) has 2-3 predicted COs per chromosome while  $\sim 18$  chiasmata are usually observed on a spread. This suggests that the number of COs is slightly underestimated using cytology in this species. In wheat, the number of COs determined according to genetic map is 3.60 COs per bivalent [109] which corresponds to  $\sim 76$  COs for the whole genome and should reflect almost the same number of chiasmata. Earlier study observed 2.97 chiasmata per bivalent and an average of 48.3 chiasmata for the entire genome [187]. Similarly to barley, this indicates that the number of COs is largely underestimated using cytology in wheat as well.

Our observations rely on chiasmata counting of Pollen Mother Cells (PMC) that underwent a complete meiosis until metaphase I. However, the genes concerned by the deletions could have impacted the earlier phases of meiocyte development and maybe lot of cells could have died before reaching metaphase. We did not observe loss in fertility in any of the aneuploid lines we used but we will have to check the amount of pollen emitted by mutant plants and compare with wild plants. This will be an important information since mutated genes could reach an earlier control for pairing and recombination with regard to the observation of meiotic behaviour.

We did not observe any effect on chromosome pairing and recombination of the partial or total loss of the short arm of the chromosome 3B which suggests that this arm does not carry any gene involved in pairing or recombination. If this is the case, their effect should be tiny. Because the gene *Ph2* is located on the short arm of chromosome 3D [132,133] we expected that homoeologous copy on chromosome 3BS could have played a role even if *Ph2* is mainly involved in homoeologous pairing. According the results we observed, it seems that the effect of this copy is very low and could not be detected in our conditions.



Maybe the analysis of a higher number of PMCs would have made the low differences statistically significant or maybe we should have conducted the analysis of the deletion lines in a *ph1* mutant context to get rid from its major effect on homoeologous pairing. Such a material should be developed to confirm this hypothesis. Also, we cannot exclude the fact that the *Ph2-B* copy could be deleted like this is the case for some of the homoeologous copies of the *CDK2* genes at the *Ph1* locus [129,188]. However, this hypothesis will be possible to confirm only when *Ph2* will be cloned and characterized.

On the contrary, partial or total deletion of the long arm of chromosome 3B revealed more contrasted results and as soon as a terminal deletion is created, the pairing and recombination is largely disturbed. Moreover, a complete removal of the long arm increased dramatically the perturbation of the pairing and decreased recombination which suggests the presence of at least two genes located one in deletion bin 3BL7-0.63-1.00 and the second in C-3BL2-0.22. For the former bin, evaluation of wheat radiation hybrids indicated the presence of a gene (*TaDES2*) which when deleted, lead to a reduced fertility that could be due to problem in pairing [189] which would confirm our results. In addition, Dt3BS was previously described like asynaptics which implies a defect in pairing [180]. Implication of the long arm of chromosome 3B in pairing is in accordance to previous results which mentioned the effect of this region either directly in wheat [135,157,168] or with crosses with related species such as *Aegilops* [165,190]. Moreover, the homoeologous chromosome arms 3AL and 3DL carry promoters of pairing that could be homoeoalleles of the gene on 3BL [132,134,138].

The effect of the deletion of the long arm of chromosome 3B (Dt3BS) was even higher than the total deletion of chromosome 3B (N3B) with ~26 and ~35 chiasmata respectively. This may reflect the fact that some genes from the short arm of chromosome 3B act as complex with those from the long arm and regulate their activity. Once the short arm only is present the genes may regulate more strongly the homoeologous copies from chromosome arms 3AL and 3DL.



## B. Identification of candidate genes located in the two deletion bins concerned

In order to identify putative candidate genes involved in pairing and recombination, we exploited the first ordered, annotated and assembled in a pseudomolecule sequence of wheat chromosome 3B [121]. Using *in silico* assignment of genes known to be involved in meiosis in other species and RNA-seq data we identify several genes mapped in deletion bins C-3BL2-0.22 (*RTEL1*, *DUO1* & *MPK*) and 3BL7-0.63-1.00 (*MLH1*, *ZIP4*, *SYN4*, *MUS81*, *ASK2* & *CAP*) which are the two bins that had a major impact on pairing and recombination on chromosome 3B. Expression analysis indicates that chromosome-3B homoeologous copy is the most expressed for *MPK*, *MLH1* and *CAP*.

*MPK4* (Mitogen-activated protein kinase) was initially described as playing a role in regulating cortical microtubule bundling [191] but recently it was evidenced that *mpk4* mutants develop normal microspore mother cells that further fail to complete meiotic cytokinesis leading to aborted pollen grains [192]. Here we find a related MPK expressed during meiosis suggesting an implication in regulating meiosis division.

*MLH1* belongs to the DNA mismatch repair system (MMR) and is homologue to the *MutL* gene originally described in *Escherichia coli*. Three homologues of *MutL* exist in Arabidopsis (*AtMLH1*, *AtMLH3* and *AtPMS1*) and it was shown that *AtMLH1* and *AtMLH3* act as dimer during the pachytene stage of meiosis [193]. The mutation of *AtMLH1* evidenced that it is required for homologous recombination and that it also acts to limit recombination between divergent sequences [194]. In bread wheat, it was shown that *MLH1* is under the control of the *Ph1* locus [129] which prevents *MLH1* sites from becoming crossovers on paired homoeologous during meiosis [130]. In our case, *MLH1* could be a good candidate to explain the reduction in chiasmata number since it contributes to resolve the crossovers at the end of meiosis. If *MLH1* is mutated, we should thus observed less chiasma compared to the wild type.



*CAP-E1* is a functional orthologue of *SMC2* (Structural Maintenance of Chromosomes 2), a gene that belongs to the condensin complex involved in mitotic chromosome condensation and dosage compensation in yeast [195]. Two copies (*CAP-E1* and *CAP-E2*) were identified in Arabidopsis and mutant analysis revealed that the genes are expressed during meiosis and double heterozygous mutants exhibited a reduction in condensation at metaphase and anaphase I, and a mild *cut*-type phenotype at anaphase I [196]. Since chiasmata resolution requires that sister chromatids decatenate from each other along the chromosome length, when *CAP-E1* and/or *CAP-E2* are mutated, this decatenation fails and results in a lower number of chiasmata and a higher number of univalents.

Despite the fact that the homoeologous copy from chromosome 3B of these two genes are the most expressed among the three copies and that they thus constitute preferential candidates, we cannot exclude that the other genes that we located in the two deletion bins play also a role in the phenotype observed. For example *ZIP4* is involved in the formation of the synaptonemal complex [197] and *SYN4* in centromere cohesion [198]. *MUS81* is involved in class II crossovers and resolves aberrant intermediates induced during replication as well as for efficient synthesis-dependent strand annealing (SDSA) and to a smaller extent for single-strand annealing (SSA) [199]. It could thus impact homologous recombination frequency and its removal would explain the reduced number of chiasmata that we observed in the lines missing part or totality of the long arm of chromosome 3B.

Regarding genome stability or cell-cycle disruption, *DUOI* is a male germline-specific transcription factor that influences sperm-cell specification by integrating expression of male germline-genes with germ-cell-cycle progression through the G2/M phase-specific accumulation of CYCB1;1 [200] and that upregulates genes required for fertilization [201]. *RTEL1* is a DNA-helicase involved in both DNA repair and telomere stability which regulates telomere elongation and homeostasis [202,203].





*RTEL1* also maintains genomic stability by suppressing homologous recombination. In *Arabidopsis thaliana*, *RTEL1* was shown to play the same role regarding genome stability preservation as well as homologous recombination suppression [204,205]. Moreover, *RTEL1* is involved in inter- and intra-strand DNA cross-link repair and it could play a role in the destabilization of D- and T-loop structures [205]. In the Dt3BS line, we observed a lot of interlocks that were not resolved and which could be partly explained by the loss of this gene.

Moreover, we had more than 900 genes that were expressed during meiosis in the two deletion bins C-3BL2-0.22 and 3BL7-0.63-1.00. We cannot exclude that at least one of these genes that is currently unknown as playing a role in meiosis can be responsible of the observed phenotype. A more detailed study is needed to identify those for which the chromosome-3B homoeologous copy is the most expressed. Then mutants for these genes will have to be analysed individually and the resources developed for mutant identification (<http://www.wheat-tiling.com/>) will be of main importance to accelerate this process. Finally, all the genes that are located in the two bins may also act in complexes or may regulate each other which will largely complicate the identification of those having the largest impact on the phenotype we observed.



## 5. Conclusion

Our study allowed the identification of at least two genes impacting chiasmata number and located distally (3BL7-0.63-1.00) and proximally(C-3BL2-0.22) on the long arm of chromosome 3B of bread wheat. Formation of bivalents at meiosis results from three major processes: chromosome pairing itself which is the interaction between homologous chromosomes resulting in their alignment; synapsis which is the formation of the synaptonemal complex between homologues; crossovers which occur at the end of prophase I. Meiotic behavior of our mutants needs to be deciphered to better understand which part of the process is impacted and to better focus on the appropriate candidates. The recent genomic resources that are now available for wheat gave us the opportunity to find putative candidate genes. They now need to be studied in more details by searching for mutants in TILLING populations or by developing transgenic plants (knock-out or over-expression) and evaluate pairing and recombination behavior of the mutants.

## Acknowledgements

We are grateful to B Gill and J Raupp (Kansas State University, USA) for providing the deletion lines. We would like to thank our colleagues from greenhouses and GENTYANE for helping in preparing the material and for the support in the expression analysis. BD was funded by Meta-Programme SELGEN (project BO-DeLiRe) and Region Auvergne.



## 2. Principales conclusions

L'étude cytologique préliminaire menée dans l'équipe d'accueil sur les lignées de délétion du chromosome 3B a montré une réduction du nombre de chiasmas chez les lignées délétées au niveau du bras long du chromosome 3B comparées à la lignée sauvage Chinese Spring. Cette réduction semble liée à une diminution du nombre de bivalents anneaux et une augmentation des univalents et des bivalents droits. Ce phénomène est prévalent chez deux lignées : la lignée 3BL7-0.63 qui présente une délétion de la partie distale du bras long chromosome 3B et la lignée ditélosomique 3BS (Dt3BS) qui présente une délétion totale de ce même bras. Cela suggère l'existence d'au moins deux locus sur le chromosome 3B qui affectent le nombre de chiasmas.

Par la suite, l'accès à la séquence du chromosome 3B ainsi que l'utilisation de données RNA-seq a permis la cartographie *in silico* de plusieurs gènes impliqués dans le déroulement de la méiose comme une MPK, CAP et MLH1 et présents dans les différentes délétions. Ces différents candidats doivent maintenant être validés en étudiant le comportement méiotique des mutants pour la copie du chromosome 3B.



### 3. Perspectives

Ce type d'étude cytogénétique est actuellement essentiel pour analyser le comportement méiotique des mutants des gènes méiotiques. Il se rapproche du crible de mutants effectué sur *Arabidopsis* afin de découvrir les gènes impliqués dans la méiose. Compte tenu des tailles des délétions, il n'est pas possible de déterminer clairement quels gènes sont responsables des phénotypes. Nous avons focalisé sur des gènes connus comme étant impliqués dans la méiose mais il n'est pas exclu que le phénotype observé soit dû à d'autres gènes inconnus actuellement et différenciellement exprimés pendant la méiose. Néanmoins, l'information de séquence nouvellement disponible permet d'établir des candidats qui peuvent être testés à l'aide de collections de mutants ou de mutants ciblés (CRISPR-CAS9).

La recherche des gènes impliqués dans la méiose et le nombre de chiasmas est d'une grande importance chez le blé dans la mesure où chez cette espèce, la recombinaison est faible et mériterait d'être améliorée. La connaissance des gènes méiotiques et de la mécanistique des plantes modèles comme *Arabidopsis* nous a beaucoup appris sur la manière dont la méiose se déroule chez le blé et sur les potentiels leviers qui peuvent être utilisés pour contrôler/favoriser/réduire la recombinaison. Néanmoins, la transposition vers une espèce aussi complexe que le blé reste à établir.





### III. Partie II – Analyse de la recombinaison au niveau du génome entier

---

#### 1. Introduction

La complexité du génome du blé tendre a longtemps été un frein au développement de ressources génomiques et en particulier des marqueurs moléculaires pour couvrir les 17 Gb du génome afin de réaliser des analyses de génétique quantitative, de la sélection assistée par marqueur (MAS), etc...

Avec les nouvelles technologies de séquençage, le développement de SNP a explosé ces dernières années ouvrant de nouvelles perspectives pour les analyses de la structure et de l'évolution du génome[206]. Dans cet article (soumis pour publication à Molecular Plants), j'ai réalisé la totalité des cartes génétiques après automatisation de la création des fichiers de cartographie. Pour des raisons de confidentialité, seules les données concernant ~280 000 SNP ont été publiées alors que j'ai traité plus de 420 000 marqueurs (voir partie 3.A). Ces données ont été aussi l'occasion de découvrir des biais de ségrégation dans la population Chinese Spring X Renan, de comparer la relation entre distances physique et génétique de tous les chromosomes du blé tendre à l'aide des nouvelles pseudo-molécules NRGeneV3, de comparer la recombinaison entre trois chromosomes d'un même groupe d'homéologie (3A 3B et 3D), et d'appréhender la recombinaison à l'échelle du génome entier.



# High throughput SNP discovery and genotyping in hexaploid wheat

---

Hélène Rimbart<sup>a,2</sup>, Benoît Darrier<sup>a,2</sup>, Julien Navarro<sup>a</sup>, Jonathan Kitt<sup>a</sup>, Frédéric Choulet<sup>a</sup>, Magalie Leveugle<sup>b</sup>, Jorge Duarte<sup>b</sup>, Nathalie Rivière<sup>b,3</sup>, The International Wheat Genome Sequencing Consortium<sup>c</sup>, The BreedWheat Consortium<sup>d</sup>, Alessandro Davassi<sup>e</sup>, François Balfourier<sup>a</sup>, Marie-Christine Le Paslier<sup>f</sup>, Aurélie Bérard<sup>f</sup>, Dominique Brunel<sup>f</sup>, Catherine Feuillet<sup>a,3</sup>, Charles Poncet<sup>a</sup>, Pierre Sourdille<sup>a</sup>, Etienne Paux<sup>a,1</sup>

a GDEC, INRA, Université Blaise Pascal, 63039 Clermont-Ferrand, France

b Biogemma, 63720 Chappes, France

c IWGSC, Eversole Associates, Bethesda, MD 20816, USA

d BreedWheat, 63039 Clermont-Ferrand, France

e Affymetrix, High Wycombe, HP10 0HH, UK

f EPGV US 1279, INRA, CEA, IG-CNG, Université Paris-Saclay, 91000 Evry, France

1 To whom correspondence should be addressed. [etienne.paux@inra.fr](mailto:etienne.paux@inra.fr)

2 These authors contributed equally to this work.

3 Current address: Limagrain, 63720 Chappes, France

4 Current address: Bayer CropScience, Morrisville, NC 27560, USA

Running title: SNP discovery and genotyping in wheat



## 1. Abstract

Because of their abundance and their amenability to high-throughput genotyping techniques, Single Nucleotide Polymorphisms (SNPs) are powerful tools for efficient genetics and genomics studies, including characterization of genetic resources, genome-wide association studies and genomic selection. While the size and complexity of the bread wheat (*Triticum aestivum* L.) genome has long hampered SNP discovery, the advent of next-generation sequencing opened new perspectives for the detection of polymorphisms and unlocked the development of several millions of SNPs as well as high-throughput genotyping arrays for hexaploid wheat. However, most of these initiatives targeted the coding fraction, leaving almost 98% of the wheat genome largely unexploited. Here we report on the use of whole-genome resequencing data from eight wheat lines for the discovery of 3.3 million SNPs distributed in the genic and as well as the repetitive and non-repetitive intergenic fraction of the wheat genome. We also describe the development of the TaBW280K high-throughput genotyping array containing 280,226 SNPs and its use to characterize a set of 96 wheat accessions and to construct a dense genetic map comprising 53,287 markers. The results described here will provide the wheat community with powerful tools for both basic and applied research.

Keywords: wheat, SNP discovery, genotyping, genetic diversity, presence-absence variations, genetic map



## 2. Introduction

Because they are the most abundant type of polymorphism in plant and animal genomes and because they are amenable to high-throughput, cost effective genotyping technologies, Single Nucleotide Polymorphisms (SNPs) have been adopted as the markers of choice in genetics. In the past years, they have been widely used for various applications, including genome-wide association studies, characterization of genetic resources, marker-assisted breeding and genomic selection [207,208]. The power of these different approaches relies heavily on marker density and on the ability to assay thousands of SNPs in parallel. Compared to other types of markers such as simple sequence repeats (SSRs) or Diversity Array Technology (DArT) markers, SNP discovery relies on the comparison of homologous sequences between genotypes to identify variations at the sequence level [209]. If the advent of next generation sequencing systems opened the way to whole genome resequencing of several small to medium genome plant species for SNP discovery [210–214] the size of the wheat genome has long hampered such approaches. As a result, most of the SNP discovery initiatives that have been conducted to date relied on complexity reduction approaches. For example, [215] used an exome capture array to target ~57 Mb of coding sequences in 43 bread wheat accessions and wheat relatives and discovered 921,705 putative varietal SNPs. Similarly, a wheat exome capture targeting 107 Mb of non-redundant genic regions was used by [216] to mine for SNPs in a panel of 62 wheat lines. Eventually, ~1.57 million SNPs were identified. In 2014, RNA-seq on a set of 19 bread wheat accessions led to the discovery of 67,686 variants [217]. Also, genotyping-by-sequencing has been applied to wheat.





Despite the fact that this technique has the potential to sample a higher fraction of the genome and especially intergenic regions that are not targeted by exome capture, the large amount of missing data limits SNP discovery. Recently, this approach was used by [216] on a set of 62 diverse hexaploid lines and [218] on the NIAS Japanese Wheat Core Collection consisting of 96 lines. In both studies, the number of SNPs was in the range of 200,000 to 300,000. A noticeable exception to complexity reduction approaches is the work conducted by [219] who used whole-genome resequencing data from 16 wheat accessions to perform SNP discovery. In this study, more than four million SNPs were discovered. However, a focus was made only on group 7 chromosomes and consequently, similar data have never been produced at the whole-genome level.

Concomitantly to SNP discovery, several technologies have been implemented for SNP genotyping, from low-throughput monoplex to ultra-high-throughput highly-multiplexed assays. Low- to medium-throughput technologies include padlock probes

[220], Applied Biosystems SNPLex [221] and SNaPshot Multiplex System [222], Illumina VeraCode [222,223], LGC Genomics KASPar assays [224,225], and Sequenom Mass Assay [221]. More recently, with the availability of large amount of SNP markers, high throughput assays have been successfully developed in wheat. In 2013, [226] developed an Illumina iSelect array containing ~9,000 gene-associated SNPs and used it to genotype 2,994 hexaploid wheat accessions. The year after, the same technology was used to design a 90K SNP array and characterize the wheat diversity in a set of almost 2,500 accessions [217]. In 2015, [215] reported on the development of an Affymetrix Axiom 820K SNP array and its utilization to genotype 475 hexaploid wheat and wheat relative accessions. All these high-throughput arrays contain mainly gene-derived SNPs.



However, bread wheat being an allohexaploid species, genes are present in multiple copies. In addition, the wheat genome has undergone more interchromosomal duplications than related grasses, with 20 to 30% of genes being duplicated [121,122,143]. As a result, genotyping is complicated by the presence of homoeologous and paralogous loci [217,226,227]. So far, the intergenic fraction of the wheat genome has been poorly exploited for SNP discovery. Nevertheless, while genes account for 1-2% of the genome, non-coding regions represent 98-99%, with 10-15% being low-copy sequences. Also, the repetitive fraction has been shown to be a great source of SNPs through the use of Insertion Site-Based Polymorphism (ISBP) markers [222,225]. Here we report on the exploitation of the whole wheat genome sequence for SNP development and genotyping. Using whole-genome resequencing data from eight wheat accessions, we discovered more than three million genome-wide SNPs and mined for single-copy loci to design a high-throughput genotyping array containing 280,226 SNPs. Out of them, 68.5% were converted, with almost one half producing diploid-like clusters, therefore demonstrating the efficacy of our design strategy. We also present the characterization of a 96-accession panel, as well as the construction of a dense genetic map comprising approximately 53,000 loci.

**Table 1.** Summary of the number of SNPs

	ISBP-derived	Intergenic	Exonic	Intronic	Total
A-genome	390,462	866,599	38,485	61,859	1,357,405
1A	60,774	131,481	5,851	9,367	207,473
2A	75,889	153,026	7,058	11,423	247,396
3A	40,179	98,901	4,433	7,028	150,541
4A	36,467	90,892	4,464	6,241	138,064
5A	56,534	118,432	4,493	8,575	188,034
6A	54,744	113,872	6,429	9,272	184,317
7A	65,875	159,995	5,757	9,953	241,580
B-genome	469,964	1,021,845	44,266	71,207	1,607,282
1B	59,926	136,537	5,517	9,600	211,580
2B	81,977	168,903	9,864	14,169	274,913
3B	96,264	213,332	8,219	13,499	331,314
4B	32,557	62,266	3,055	4,920	102,798
5B	75,310	163,224	8,627	13,169	260,330
6B	58,675	136,036	5,071	8,371	208,153
7B	65,255	141,547	3,913	7,479	218,194
D-genome	92,876	205,378	11,451	15,455	325,160
1D	15,977	38,373	2,271	2,874	59,495
2D	17,933	38,924	2,737	4,257	63,851
3D	9,846	23,035	921	1,280	35,082
4D	8,893	14,271	511	620	24,295
5D	12,343	26,679	1,639	2,079	42,740
6D	11,065	26,582	1,664	2,072	41,383
7D	16,819	37,514	1,708	2,273	58,314
Total	953,302	2,093,822	94,202	148,521	3,289,847

### 3. Results

#### A. De novo SNP discovery in the hexaploid wheat genome

Whole genome resequencing data from eight wheat cultivars (Premio, Renan, Robigus, Xi19, Volcani, Westonia, Xiaoyan54 and Yitpi) were used to mine for SNPs in the hexaploid wheat genome. To this aim, reads were mapped to the recently released chromosome-based draft genome sequence [143]. For SNPs located in low-copy regions, the repeat-masked contigs were used as a reference. This dataset comprises 4,497,443 contigs covering approximately 2.1 Gb of unmasked sequences. For SNPs located in the repetitive fraction of the genome, ISBP markers were designed from contigs [222]. A total of 6,816,965 ISBPs were identified of which 1,003,684 were high confidence and 2,837,656 were medium confidence markers. High and medium-confidence ISBPs were subjected to clustering to discard non-unique regions. Eventually, 2,452,835 unique ISBPs covering ~523 Mb were used as a reference sequence for ISBP-derived SNP discovery.

SNP calling and filtering predicted 3,289,847 intervarietal SNPs, of which 2,336,545 were located in low-copy sequences and 953,302 in ISBPs (Table 1 and Supplemental Table S1). Strong differences were observed in terms of SNP proportion between homoeologous genomes, with the B-genome being the more polymorphic (1,607,282 SNPs; 49%), followed by the A-genome (1,357,405; 41%) then the D-genome (325,160; 10%). Within-subgenome differences were also observed. For example, in the B-genome, SNP density ranges from 125 SNPs / Mb on 4B to 334 SNPs / Mb on 3B. Finally, the SNP density was also found to vary greatly between homoeologous groups, the group 4 representing only 8% of the whole SNPs, whereas the other groups represent 15% on average.

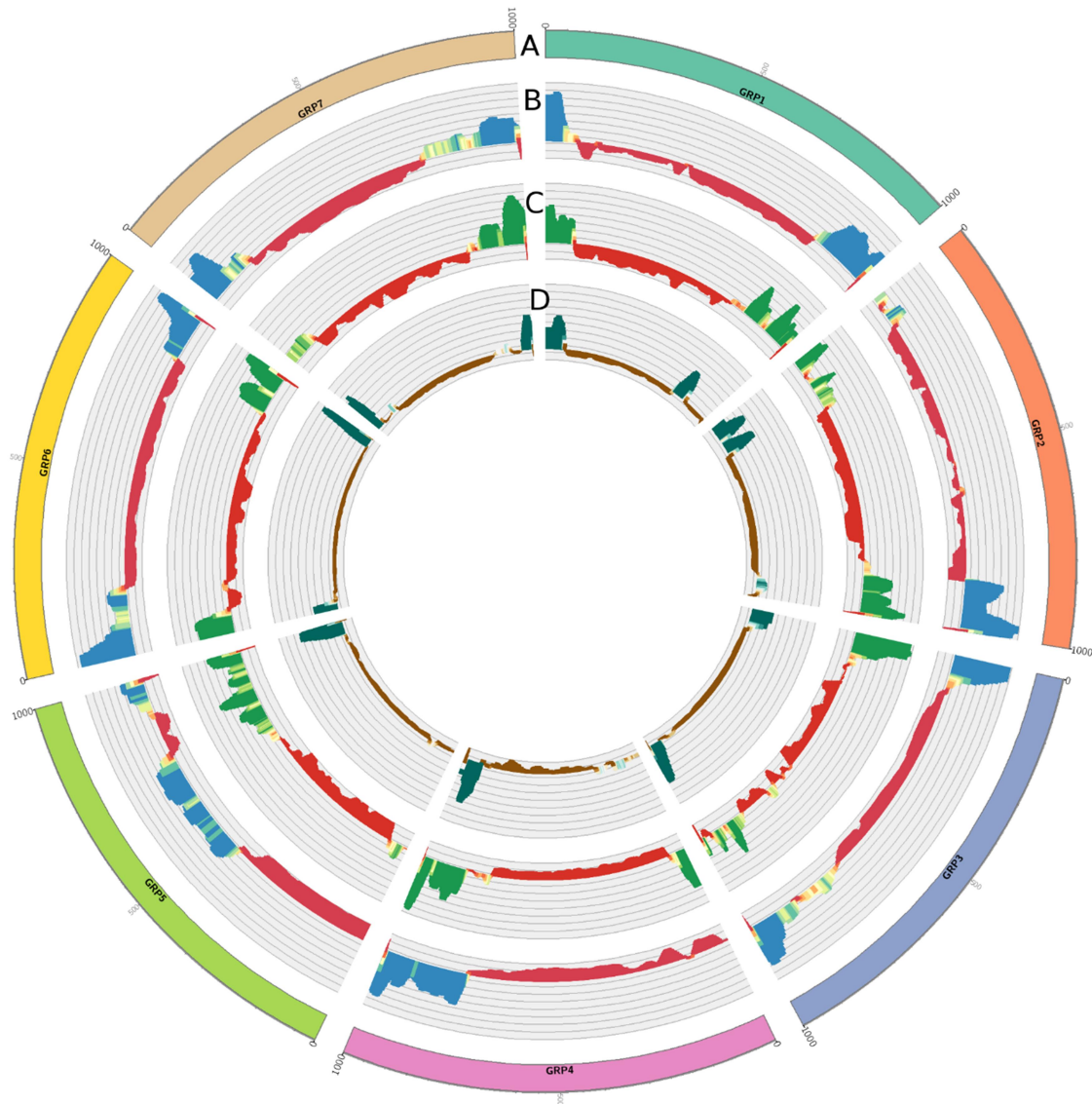


In addition to a reduced diversity level, the D-genome also exhibits a higher percentage of private SNPs, *i.e.* polymorphisms observed in only one line (on average, 29% for the D-genome *vs.* 21% for the A- and B-genomes).

The majority of SNPs (68%) corresponded to transition (Ts) and 32% to transversion (Tv), leading to an average Ts/Tv ratio of 2.12, a value that is similar to what has been reported previously in wheat [215,217,225]. This ratio was found to be higher in ISBPs (2.28) than in low copy regions (2.06), probably as a result of the high methylation level in TEs that leads to an increase in mutation frequency at deaminated sites [222,225,228,229]. Gene models from the chromosome-based draft genome sequence were used to assign polymorphisms found in non-repetitive sequences to either genic or intergenic compartments using SNPeff [230]. Out of the 2,336,545 SNPs, 2,093,822 were found in intergenic regions. The remaining 242,723 were found in 45,114 out of the 99,386 predicted protein-coding genes. Of these, 94,202 were located in exons and 148,521 in introns. In coding regions, 45% of SNPs corresponded to synonymous substitutions while 55% were non-synonymous, with 1.4% of non-sense mutations. No differences were observed between genomes. The functional effect of non-synonymous SNPs was then predicted using the Grantham matrix [231]. Consistent with previous estimates [143], 35% of non-synonymous SNPs were conservative, 44% moderately conservative, 14% moderately radical and 7% radical.

In order to have an overview of the genomic organization of the nucleotidic variations in the wheat genome, SNPs were mapped onto the POPSEQ-ordered whole-genome draft assembly [143,232].



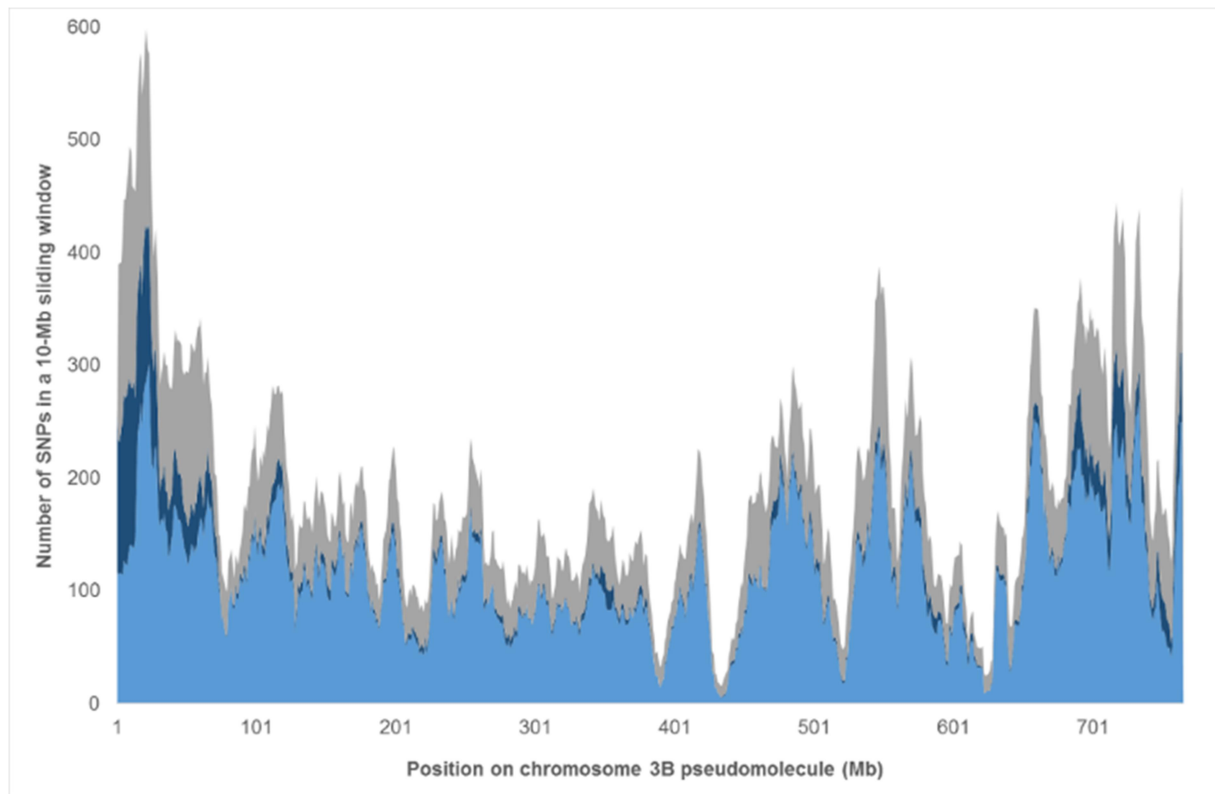


**Figure 1.** Genome-wide distribution of SNPs along wheat chromosomes. Circos map displays (A) homoeologous groups, as well as SNP density along (B) A-genome, (C) B-genome and (D) D-genome chromosomes. SNP density is expressed as a Z-score of the density in a 10-Mb sliding window (step 1 Mb).

Out of the 3,289,847 polymorphisms, 2,310,481 (70.2%) were present in this assembly. This includes 1,483,013 intergenic (71%), 185,863 genic (77%) and 641,815 ISBP-derived SNPs (67%). The higher percentage observed for genic SNPs illustrates the fact that the POPSEQ-ordered whole-genome draft assembly is enriched in gene-containing contigs. The number of mapped SNPs is well correlated with the total number of SNPs ( $R=0.98$ ;  $p=0$ ). As expected, SNPs were generally more abundant in the distal parts of most chromosomes (Figure 1). However, differences were observed between chromosomes. For example, the short arm of chromosomes 4A, 4D, 5A and 5D appeared to be less polymorphic than the long one. In addition, on 5A and 5B, polymorphisms were found to spread over a larger portion of the chromosome.

### **B. Development of a high density SNP genotyping array for wheat**

A high-density Affymetrix Axiom genotyping array was designed using a subset of the 3.3 million SNPs. For the selection of SNPs, several criteria were applied. First, only SNPs found in at least two out of the eight lines and polymorphic in European germplasm were selected. Second, the number of SNPs per chromosome arm was chosen according to arm size. However, because of the natural lack of polymorphism in the D-genome, we decided to double our genotyping array in D-genome-originating SNPs. Third, SNP context sequences were aligned against contigs from the chromosome-based draft sequence of the wheat genome [143] and loci having multiple hits in the genome were discarded. Finally, priority was given to intergenic SNPs, since several genic SNP-based arrays have already been developed recently [215,217,226].

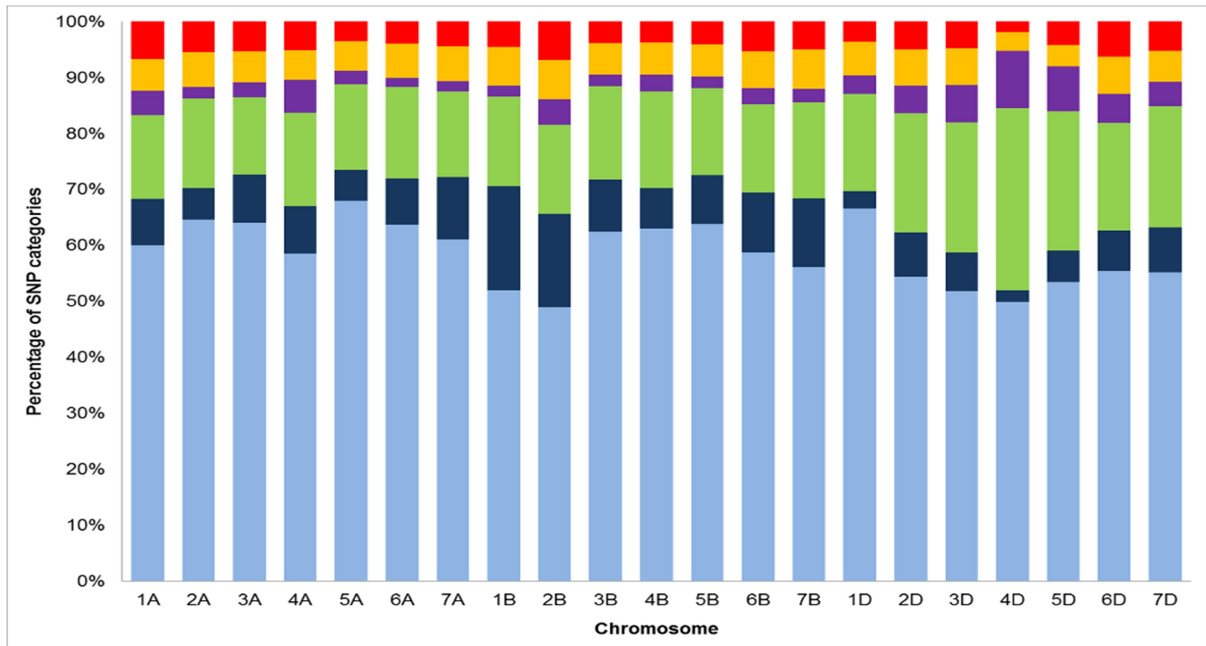


**Figure 2.** Distribution of the TaBW280K SNPs along the wheat chromosome 3B. Densities of Polymorphic High Resolution SNPs (light blue), Off-Target Variants (dark blue) and non-converted (grey) SNPs are computed in a 10-Mb sliding window (step 1 Mb). The X-axis represents the position on chromosome 3B pseudomolecule (in Mb).

Eventually, our Axiom genotyping array (hereafter referred to as the TaBW280K array) comprised 280,226 SNPs, including 225,596 intergenic and 54,280 genic SNPs (Supplemental Table S2). The percentage of SNPs for the A-, B- and D- genome was 39%, 42% and 19%, respectively. Regardless of the genome, the number of SNPs per chromosome arm was highly correlated with the arm size ( $R>0.9$ ;  $p<1E-6$ ). The SNP density was one marker every 52 kb on the A-genome, one every 53 kb on the B-genome and one every 92 kb on the D-genome. These 280,226 SNPs originated from 117,799 independent IWGSC contigs and covered 19,085 predicted genes.

In the absence of a whole genome reference sequence, SNPs were selected randomly on each chromosome arm. However, since good coverage of the genome is an important feature of genotyping arrays, we assessed the quality of our coverage by studying physical distribution of selected SNPs along chromosome 3B, the only chromosome for which a reference sequence was available [121]. Out of the 280,226 SNPs, 18,745 were assigned to this chromosome. By aligning SNP context sequences to the 3B pseudomolecule, 15,123 were unambiguously mapped to a single position. The average density was of 192 SNPs / 10 Mb and the median density, 169 SNPs / 10 Mb, ranging from 16 to 592 (Figure 2). Only 7.7% of inter-SNP distances were larger than 200 kb, the largest region without any single SNP being 2.56 Mb-long. Two main regions of low SNP density were observed on the long arm of chromosome 3B (approx. 430-440 Mb and 620-630 Mb).

Performance of the TaBW280K genotyping array was examined by genotyping a set of 96 wheat accessions comprising 13 European elite varieties as well as 83 accessions from a core collection representing the world-wide diversity [233] (Table S3).



**Figure 3.** Proportion of TaBW280K SNP clustering categories. For each chromosomes, the percentage of each category is displayed: Polymorphic High Resolution (light blue), Off-Target Variants (dark blue), Monomorphic High Resolution (green), No Minor Homozygous (purple), Call Rate Below Threshold (yellow) and Others (red).

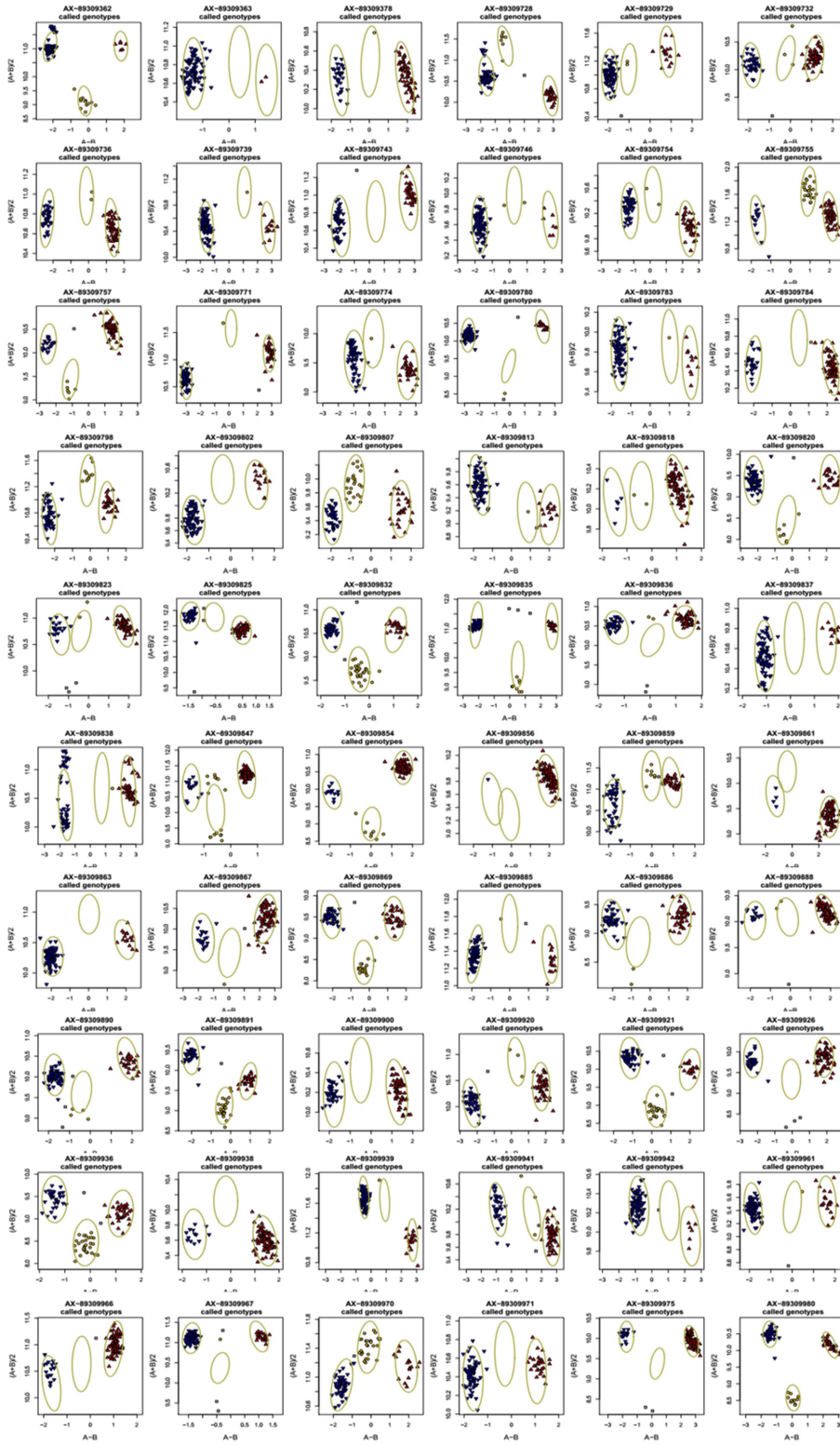
**Table 2.** Summary of the number of SNPs per clustering category and clustering metrics.

	Total	Genic	Intergenic
Total	280,226	54,280	225,946
PHR	165,885	38,276	127,609
OTV	26,105	4,182	21,923
MHR	48,204	6,233	41,971
NMH	9,730	361	9,369
CRBT	16,732	2,869	13,863
Other	13,570	2,359	11,211
HomRO > 0.3	148,423	29,807	118,616
HomFLD > 10	100,138	21,873	78,265
Diploidized	92,419	19,158	73,261

SNPs were classified in six main categories according to cluster patterns produced by the Affymetrix software: Polymorphic High Resolution (PHR; 165,885; 59%), Off-Target Variants (OTV; 26,105; 9%), Monomorphic High Resolution (MHR; 48,204; 17%), No Minor Homozygous (NMH; 9,370; 4%), Call Rate Below Threshold (CRBT; 16,732; 6%) and Others (13,570; 5%) (Table 2; Supplemental Table S2). Differences were observed between chromosomes (Figure 3). For example, chromosomes 1B and 2B displayed the highest proportion of OTVs, while 4D had the highest proportion on MNH. Differences were also observed between homoeologous genomes, with the B-genome having more OTVs and the D-genome having more MNH.

For subsequent analyses, only PHR and OTV SNPs were considered as converted SNPs, *i.e.* polymorphic SNPs that can be efficiently scored on the array. This represented 191,990 SNPs, thus an overall conversion rate of 69%. This conversion rate differed according to SNPs, with 78% of genic and 66% of intergenic SNPs being converted. The number of converted SNPs per chromosome was highly correlated with the overall number of SNPs present on the chip ( $R=0.99$ ;  $p=0$ ). The converted SNP density was one marker every 74, 79 and 158 kb on the A-, B- and D-genomes, respectively. The absence of bias in the conversion rate was confirmed at the sequence level on chromosome 3B where the number of converted SNPs in a 10-Mb window was highly correlated with the overall number of SNPs in the same window ( $R=0.98$ ;  $p=0$ ) (Figure 2).

The overall heterozygosity rate was found to be 1.5%, as expected for pure homozygous lines with limited residual heterozygosity rate. The percentage of missing data per genotype ranged from 0.4 to 5.1%, with 88% of the lines having less than 2% missing data. This can be extrapolated as a 91% “call rate” over the whole analysis. Finally, the percentage of total genotypes returned across all markers, defined as the “completeness”, was 99%, with 1% missing data.



**Figure S1.** Examples of diploidized ( $\text{HomRO} > 0.3$  and high  $\text{HomFLD} > 10$ ) SNPs. A-B: Contrast =  $\log_2[(\#A \times 100 + 100) / (\#B \times 100 + 100)]$ ; (A+B)/2: Size =  $[\log_2(\#A \times 100 + 100) + \log_2(\#B \times 100 + 100)] / 2$ .

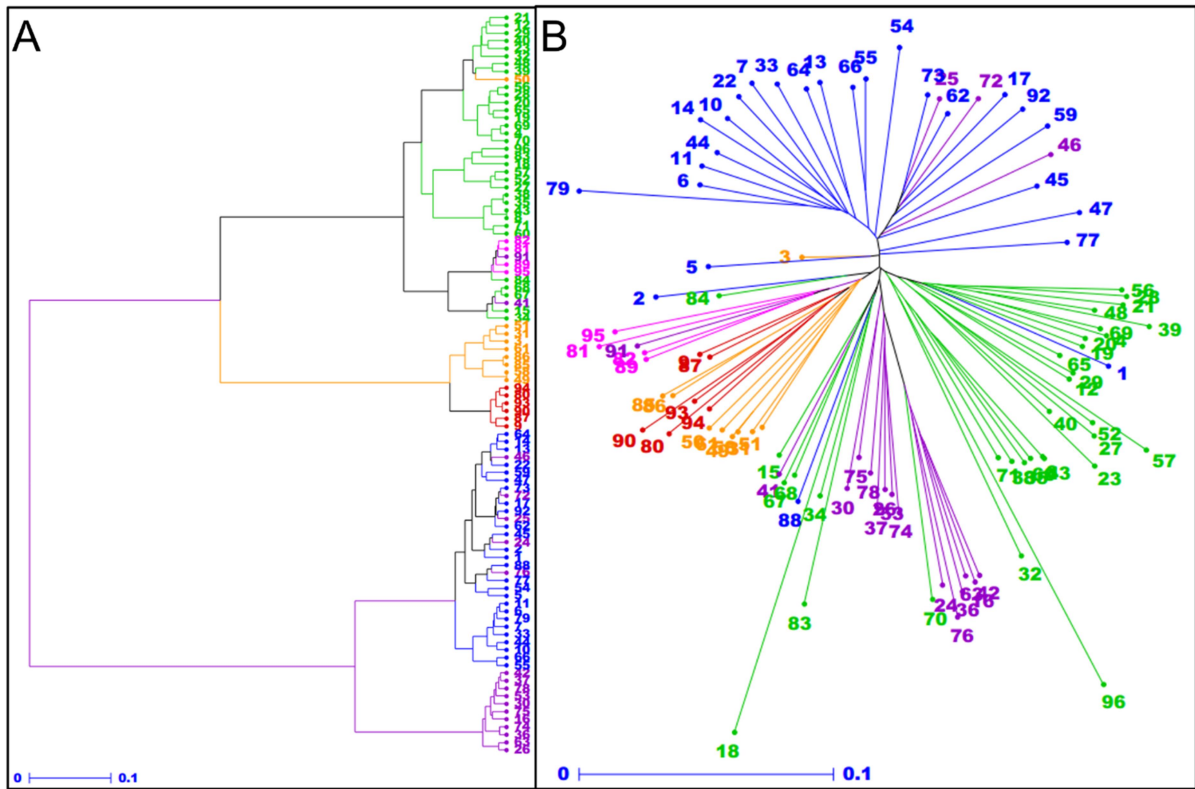
Two other metrics can be used to check probeset quality on Axiom arrays. The first one is the Homozygote ratio Offset (HomRO) that defines the location in the contrast dimension (X-axis) of the homozygous genotype clusters. Positive HomRO values indicate that AA and BB clusters are located on each side of zero. Negative values indicate that both clusters are on the same side of zero. Theoretically, in polyploid species where several subgenomes contribute to allele dosage, HomRO values tend to be low. In tetraploid species, HomRO is expected to be close to 0 and below 0.3 and in hexaploid species, below 0 [234]. In our experiment, 148,423 probesets (77% of converted probesets) had a HomRO  $> 0.3$ . The second QC metric is the Homozygote Fisher's Linear Discriminant (HomFLD) which is a measurement of the cluster quality of a SNP. HomFLD examines the distance between the two homozygous clusters as well as the variance across clusters. High FLD values indicate high-quality clusters with well-separated centers and narrow distribution. Classically, for polyploid species, Affymetrix recommends a HomFLD value greater than 3.6. By manually examining SNP clusters, we found that a HomFLD value greater than 10 was a good, yet very conservative, indicator of high quality clusters. In our dataset, 100,138 probesets (52%) had HomFLD values greater than 10. By combining diploid-like HomRO ( $>0.3$ ) and high HomFLD ( $>10$ ) values, we identified a set of 92,419 SNPs (48%) that can be considered as diploidized (see Supplemental Figure S1 for examples). The percentage of diploidized SNPs slightly differs between genic and intergenic, with 45% and 49%, respectively. While the percentage of probesets having a HomFLD  $> 10$  is similar for both types of SNPs, the main difference comes from the polyploidized SNPs (HomRO  $< 0.3$ ), 30% of genic SNPs and 21% of intergenic SNPs.





The average Polymorphism Index Content among the 96 lines was 0.36, corresponding to an approximate 25 vs. 75% distribution. This value is comparable to previously reported PIC values in wheat [224,225]. It is worth noting that 50% of the SNPs had a PIC value between 0.4 and 0.5 and only 3,165 (2%) were found in one single genotype.

As suggested by [235], OTVs can be used to detect presence / absence variations (PAVs). In our experiment, 26,105 SNPs were classified as OTVs. The average percentage of converted SNPs corresponding to OTVs per line was 1.5%, ranging from 0.1% for CS to 2.7% for Equinox. In the top ten lines showing the highest OTV rate (>2.0%) were eight lines in which more than 40% of the OTVs were located on the short arm of chromosome 1B (1BS), while the average rate on 1BS for the whole panel was 0.7%. At least six out of these eight lines are known to carry the 1RS.1BL translocation. Another line was BlueBoy in which almost 40% of OTVs were located on chromosome 2B (vs. 11.4% for the other lines on average). Finally, not surprisingly, the last one of the top 10 lines was W7984, a synthetic wheat formed by hybridizing a tetraploid wheat *Triticum turgidum* L. subsp. *durum* var 'Altar 84' (AABB genotype) with the diploid goat grass *Ae. tauschii* (219; CIGM86.940) (DD genotype) [236]. On chromosome 3B, OTV were mainly found in the fast-evolving distal R1 and R3 regions identified by [121] (Figure 2).



**Figure 4.** Phylogenetic relationships between 96 wheat accessions. Ward dendrogram of the 96 wheat accessions. (A) Ward dendrogram showing phylogenetic relationships between wheat accessions revealed by PHR and OTV SNPs. (B) Neighbour-joining tree showing phylogenetic relationships between wheat accessions (Supplemental Table S3) revealed by OTV SNPs only. Accessions are numbered following the Supplemental Table S3 and colored according to their geographical origins: Asia (orange), Black Sea (pink), Persia (red), Western Europe (blue), Eastern Europe (purple) and Mediterranean (green).

### C. Phylogenetic analysis of wheat lines

Genotyping data of the 96 wheat accessions with 191,900 PHR and OTV SNPs were used to describe diversity within the panel. Figure 4A illustrates the Ward dendrogram of the 96 wheat accessions. In a previous study on a 367-individual core collection capturing 98% of the worldwide diversity, [142] performed a population structure analysis that led to five groups of accessions which can be related to their geographical origins: Western Europe (WE cluster), Eastern Europe (EE), Mediterranean (Med), Asia (AS) and Nepal (NP). In the present study, the AS cluster was split into three main groups. The first group corresponds to Asian lines from Japan, China, Korea and India. It is worth noting that the NP group was found to be included in this cluster, close to Indian lines. A second group includes accessions from Turkmenistan, Afghanistan, Tajikistan, Pakistan and Armenia, roughly corresponding to the former Persia. The third group comprises lines from Turkey, Georgy, Azerbaijan and Russia and could be considered as the 'Black Sea' group. This group was closer from Med cluster than from the AS one. Some lines from Canada, Finland, Algeria and Greece were separated from the Med cluster and grouped together, close to the 'Black Sea' group. Finally, most of the elite varieties were clustered together.

In addition, a set of 5,486 OTVs without any single missing data was used to construct a PAV-based phylogenetic tree (Figure 4B). The geographical groups identified with the whole set of converted SNPs was largely recovered, suggesting that structural variations of the wheat genome, and more precisely PAVs, reflect geographical origins of lines. However it is worth noting that the different geographical groups were not clustered in the same way. Indeed, while the Black Sea cluster was found to be close to the Med one, it appears more closely related to the Asia and Persia clusters when considering only OTVs. Also, the EE lines were clustered with the Med and Asia groups.

**Table 3.** Description of the TaBW280K-based Chinese Spring x Renan genetic map

	Number of SNPs	Map length (cM)	Number of unique genetic positions	Number of non-redundant SNPs	Bin length (cM)
A-genome	21,718	1,019	2,080	670	0.50
1A	3,394	140	368	106	0.38
2A	3,609	94	246	81	0.38
3A	3,126	150	281	92	0.54
4A	3,074	157	255	101	0.62
5A	2,654	185	370	125	0.50
6A	2,553	123	202	66	0.61
7A	3,308	169	358	99	0.47
B-genome	21,916	877	1,995	615	0.45
1B	3,484	112	342	99	0.33
2B	2,669	108	254	72	0.43
3B	3,841	155	361	111	0.43
4B	2,233	104	210	75	0.50
5B	3,073	156	308	93	0.50
6B	3,814	124	249	67	0.50
7B	2,802	119	271	98	0.44
D-genome	9,653	1,218	1,632	937	0.75
1D	1,421	156	185	113	0.84
2D	1,477	144	182	98	0.79
3D	1,540	191	270	152	0.71
4D	1,049	144	256	146	0.56
5D	1,102	220	251	161	0.88
6D	1,601	168	218	118	0.77
7D	1,463	195	270	149	0.72
Total	53,287	3,114	5,707	2,222	0.57

Finally, five out of the six lines harboring the 1RS.1BL translocation mentioned previously were found to cluster together whereas they were dispersed along the whole tree when considering all PHR and OTV markers.

#### **D. Construction of a high-density SNP genetic mapping**

The TaBW280K SNP array was used to genotype 430 Single Seed Descent (SSD) individuals derived from a cross between Chinese Spring and Renan (CsRe) [121]. Out of the 280,226 probesets, 132,549 were found to be polymorphic between the two parental lines and PHR on the population. In order to build a robust and highly reliable map, a set of 55,872 diploidized SNPs ( $\text{HomFLD} > 10$  and  $\text{HomRo} > 0.3$ ) was used for mapping. Eventually, 53,287 (95.4%) SNPs were genetically mapped in 21 linkage groups corresponding to the 21 chromosomes of bread wheat, with no unlinked markers (Table 3; Supplemental Table S4). The D-genome was the less covered with 18% of the mapped markers while the A- and B-genomes were similarly covered with 41% of mapped markers. The most populated chromosome was 3B (3,841) and the least populated, chromosome 4D (1,049). Overall, the number of mapped SNPs was found to be highly correlated with the total number of PHR markers per chromosome ( $R = 0.94$ ,  $p < 0.00001$ ).

The genetic map covers 3,114 cM, with an average density of one SNP every 0.06 cM. The D-chromosomes had the longest genetic maps, with an average of 174 cM and a cumulative length of 1,218 cM, followed by the A-chromosomes (mean length = 146 cM; cumulative length = 1,019 cM) and the B-chromosomes (mean = 125 cM; cumulative = 877 cM).

Out of the 53,287 SNPs, 51,065 (96%) cosegregated with at least another marker, while 2,222 corresponded to non-redundant loci. The proportion of non-redundant SNPs was higher on the D-genome (10%) than on the A- and B-genomes (3%), as a result of the lower number of markers and higher recombination rate.



The number of unique genetic positions was 5,707 and the average size of the corresponding genetic bins was 0.57 cM, ranging from 0.45 to 0.75 cM, for the B- and D-genomes, respectively. Seventy-six percent of genetic bins were smaller than 0.3 cM and only 1% larger than 5 cM. The three largest gaps were located on chromosomes 6D (25.6 cM), 7A (12.7 cM) and 4A (12.4 cM).

Distorted loci ( $P < 0.01$ ) represented 8% of the markers. They were located on almost all chromosomes, except 4A, 5B, 5D, 6D and 7A. Important distortion (more than 10% of the markers) was observed for 2B, 2D, 4D, 5A, 5B, 6B, with 60%, 35%, 37%, 24%, 13% and 21% of the markers being distorted, respectively. Distortion was in favour of Renan alleles for chromosomes 2B, 2D, 4D, 5A and 5B, in favour of Chinese Spring alleles for chromosome 6B.

#### **4. Discussion**

By combining the recently released draft assembly of the hexaploid wheat genome [143] with whole-genome resequencing data from eight wheat accessions, we conducted a genome-wide SNP discovery. This led to the identification of almost 3.3 million polymorphisms, the largest genome-wide SNP collection ever produced in wheat so far. Not only it provides the wheat community with an unprecedented amount of markers but also it offers the possibility to investigate not only genic but also intergenic regions, including the repetitive fraction that has been shown to be a great source of SNPs for wheat genetics and breeding [222,225]. The overall proportions and distribution of SNPs between homoeologous genomes and chromosomes were quite consistent with previous studies in wheat [216,217,219,237].





However, the proportion of D-genome SNPs identified in our study was slightly lower than previously reported (10% vs. ~17%, respectively). This is likely due to the limited genetic diversity present in our SNP discovery panel. Indeed, in this work we used eight lines including four European elite varieties, in contrast to [216], for example, who performed SNP detection among 62 lines. Consistent with that is the fact that by sequencing a set of 43 bread wheat accessions and wheat relatives, [215] found an almost identical SNP proportion between the three homoeologous genomes. This hypothesis is reinforced by a previous work in which we performed SNP detection in ISBP markers in a set of ten European elite varieties and got the same 10% SNP proportion in the D-genome (unpublished data). The higher proportion of private SNPs in the D-genome reported in this study as well as by [216] also points towards this explanation and is consistent with the hypothesis of a limited number of ancestral D-genome donors involved in the hexaploidization process [238].

These newly discovered SNPs were used to design a high-throughput Axiom SNP array comprising 280,226 SNPs. While most of the other genotyping chips already available focused on genic SNPs [215,217,226], we chose to incorporate mainly intergenic polymorphisms (81% intergenic vs. 19% genic). The percentage of converted SNPs was 69% with 59% being PHRs and 9% being OTVs. Even though the converted SNP density was not even along chromosomes, few MB-seized regions were observed. Moreover, On chromosome 3B, these regions fit with the ones that we described recently in [225], suggesting that this should not result from a bias in SNP selection but rather from low polymorphism chromosomal segments that might indicate domestication or selection footprints. Recently, [215] reported on the development of a high-throughput Axiom 820K SNP array. This array was used to screen 475 accessions.



A similar percentage of converted SNPs was obtained (67%). However, the relative proportions of the main SNP categories departed strikingly from ours, with PHR and NHM accounting for 7% and 55% of all SNPs, respectively. As the genotyped accessions were pure homozygous lines and considering the low residual heterozygosity rate expected in such material, one can hypothesize that the NMH SNPs were miscalled because of the probeset hybridization with a homoeologous locus in the genome and that the heterozygous cluster should in turn be considered as a homozygous one. The low NMH rate observed in our study (4%), together with the high percentage of diploidized converted SNPs (48%) demonstrates the efficacy of our SNP selection procedure that excluded multiple-hits loci, therefore resulting in unique probesets. In addition the higher amount of intergenic diploidized SNPs compared to genic ones strongly suggests that using non-repetitive intergenic region-derived markers is an efficient way to design pseudodiploid genotyping arrays. The genome-specificity underlying this pseudodiploid behavior might also be a key factor for the efficient scoring of PAVs. Indeed, while our panel was much narrower from a genetic viewpoint than the one used by Winfield and collaborators, our array detected a much higher proportion of OTV (9% vs. 5%, respectively). One might also consider that data originating from diploidized SNPs would allow for copy number variation analyses, as demonstrated in cattle [239] and human [240,241]. Being able to score PAVs and CNVs is particularly relevant, considering the growing body of evidences of the role of structural variations in the expression of phenotypes not only in human [242,243] but also in plants [244–248].



Here, we showed that PAVs were mainly located in the distal regions of chromosome 3B, which have been described as the fast-evolving recombinogenic chromosomal parts that are enriched in nonsyntenic genes as well as in genes differentially expressed and potentially involved in wheat adaptation [122,124]. In addition, our PAV-based phylogenetic analysis, while recovering the classical geographical clustering of wheat lines, also revealed differences that might reflect a different type of diversity and evolution. Taken together, these results are of particular interest as adaptation has been associated with intraspecific structural variations [249]. A detailed analysis of chromosome segments showing differential structural variations between populations might therefore provide with an opportunity to identify genetic factors involved in adaptive traits to specific environmental factors.

The TaBW280K SNP array was proven to be an extremely efficient tool to characterize genetic resources. While clustering of accessions was mainly consistent with a previously reported structure [142], some discrepancies were observed, especially in AS cluster. Indeed, this group was split in three (Asia, Persia and Black Sea) that are highly relevant from a geographical viewpoint. This demonstrated that the high SNP density reached with our SNP array allowed for more precise and more reliable characterization of wheat accessions than what has been achieved so far. For some Med lines, we have not been able to explain the clustering close to the 'Black Sea' based on pedigrees. However, as these accessions appear to be breeding lines, one cannot exclude that 'Black Sea' material has been used to create them.



Also, even though elite varieties were clustered together as a result of the narrow genetic diversity found in the European germplasm, they were well separated in the phylogenetic tree, therefore reinforcing the idea that the TaBW280K is a powerful tool to discriminate between closely related lines.

Finally, our SNP array was used to construct a 53K SNP genetic map. All chromosomes were densely populated, with an average of 2,537 per chromosome, ranging from 1,049 to 3,841 SNPs. The vast majority (96%) of SNPs was found to cosegregate with at least one other marker. This is likely to be due to the high number of markers relative to the small number of COs per meiosis observed in plants, as well as to the relatively small size of the mapping population [16,250]. In addition, it is well known that a significant portion of wheat chromosomes are recombination-poor, leading even distant loci to cosegregate [109,121,251]. In addition, even though the D-genome contains fewer markers than the two others, it displays a higher recombination rate. Such a negative relationship between the polymorphism level and the recombination rate has already been reported in wheat and strongly suggests an impact of the sequence divergence on the occurrence of crossing-overs [76]. In the recent years, thanks to the advent of high-throughput genotyping approaches, several high density genetic maps have been produced. For example, [217] used a combination of eight doubled-haploid mapping populations to order 46,977 SNPs with an Illumina iSelect 90K SNP array.





Using the same SNP array, a 18,601-SNP genetic map was constructed from an eight-parent MAGIC population [252]. [215] generated a consensus of three different genetic maps (Avalon x Cadenza, Savannah x Rialto and W7984 x Opata) using an Axiom 820K SNP array. [253] produced a consensus map of three RIL populations comprising 28,644 GBS tags. GBS was also used to map 19,720 SNPs in the W7984 x Opata population [232]. While it comprises a number of markers that is quite comparable to other genetic maps, our map benefits from two main advantages. First, it has been generated from a single biparental cross, therefore reducing the impact of potential structural genomic variations between parents which is a challenging problem in consensus maps [253]. Second, our map is the first high-density genetic map generated from a cross involving Chinese Spring, the cultivar that has been selected by the wheat community as the reference to produce a high quality reference sequence. In this regard, our map will be extremely useful as it will allow direct and unbiased comparison between genetic and physical positions. Finally, our genetic map also revealed a number of distorted loci in the genome. While similar segregation distortions have already been reported in other crosses [252,254,255], the proportion and distribution of distorted loci differ between each population. It has been proposed that interspecific introgression might cause such segregation distortion [252,256,257]. Renan has VPM-1 in its pedigree, a line derived from crosses between *Aegilops ventricosa*, *Triticum persicum* and an hexaploid wheat line called Marne [258,259]. The distorted loci might be indicators of the presence of *Aegilops ventricosa* chromosome segments in the genome of Renan, therefore reinforcing the idea that the TaBW280k could be used to detect introgressions in the hexaploid wheat genome and could help defining more precisely the boundaries of alien fragments, as proposed by [215].



In conclusion, the large collection of SNPs as well as the high-throughput genotyping array and the dense genetic map described here provide the wheat community with powerful tools for efficient genetic and genomic studies, as well as for marker-assisted breeding. With the reference sequence of the wheat genome to come in a few months, polymorphisms will be ordered along chromosomes, therefore allowing for a more precise characterization of haplotypes, a better definition of introgression boundaries and a faster map-based cloning of genes underlying traits.



## 5. Methods

### Plant material and DNA extraction

#### A. Sequence data

Four European elite lines, namely Premio, Renan, Robigus and Xi19, were fully resequenced using Illumina HiSeq2000 sequencer with 2 x 100-bp paired end reads with a 500-bp insert size. Whole-genome Illumina paired-end resequencing data were retrieved from Bioplatforms Australia (<http://www.bioplatforms.com/wheat-sequencing/>) from two Australian (Westonia and Volcani), one Chinese (Xiaoyan54) and one Israeli (Yitpi) wheat lines. The read number and subsequent approximate sequencing depth varied between lines: 3,196,318,742 for Renan (18-fold coverage), 3,674,485,632 for Premio (16x), 1,902,221,314 for Robigus (11x), 3,218,293,718 for Xi19 (19x), 2,041,226,160 for Volcani (12x), 1,424,795,776 for Westonia (8x), 2,436,715,062 for Xiaoyan54 (14x) and 2,222,884,584 for Yitpi (13x).

#### B. Reference sequences

Contigs from the chromosome-based draft sequence of the wheat genome [143] were used as a reference for read mapping. This set consists in 41 multi-fasta files with whole chromosome shotgun sequence (CSS) contigs longer than 200-bp, for each individual chromosome arm but chromosome 3B for which both arms were sequenced at the same time. The all dataset comprises 12,087,812 contigs representing 10,363,698,897 bp.

#### C. Repeat detection and ISBP design

CSS contigs were analyzed and masked using RepeatMasker [260] with the TREP Plus library [261,262]. RepeatMasker results were analyzed with the IsbpFinder program to detect IISBP markers [261,263]. High and medium confidence ISBPs were filtered out through clustering to discard repeated sequences.



#### D. Read mapping

Repeat-masked CSS contigs and ISBP sequences were indexed using the *index* function of the BWA software (version 0.6.1-r104) [264]. Raw reads of the eight wheat cultivars were mapped without prior quality trimming using the BWA *aln* function with no mismatch allowed in the seed ( $k=0$ ), and a maximum of two mismatches in the read ( $n=2$ ). For CSS contigs, alignments were reported in the BAM/SAM format (Sam & Specification 2011) using BWA *sampe*. Mapped reads were filtered using Samtools view [264]. Only reads that were mapped in a properly paired mapped reads ( $f=2$ ) with a mapping quality greater than 20 ( $q=20$ ) were kept for further analysis. For ISBPs, BWA *samse* was used. BWA alignment files were merged and sorted using Samtools and awk command lines.

#### E. SNP discovery and annotation

Samtools mpileup was used to convert BAM alignment format into BCF format, with parameters -u to generate BCF output, -E for extended Base Alignment Quality for higher sensitivity in local realignment around short indels, -D to output per-sample depth in BCF, -I to avoid indel calling and -f the path to the reference sequence file used for mapping. BcfTools was used for SNP calling, with options -c to make the call, -v to output potential variants positions only, -g to call genotypes at variants sites, -e to make likelihood-based analysis and -N to skip sites where the reference base is "N". Finally, vcfutils.pl varFilter was used to filter out potential SNPs with a minimum depth of coverage of 5 (-d), and a minimum Root Mean Square (RMS) mapping quality of 30. SNPs were classified into four different classes. These classes were based on the homozygous vs. heterozygous nature of the SNP in the selected panel: (1) the locus showed only homozygous AA and BB alleles among all ten lines, (2) the locus showed homozygous AA and BB and heterozygous AB alleles, (3) the locus showed only homozygous AA and heterozygous AB alleles and (4) the locus was heterozygous AB in all ten lines.





Only class 1 and class 2 SNPs were used in further analyses. Downstream analyses were facilitated with the use of BEDTools [265] and VCFTools [266] packages. GTF file corresponding to the IWGSC MIPS v2.2 High Confidence Genes [143] was used to build an annotation database with snpEff build with -q gtf22 option [230]. snpEff v4.2 was run on low copy regions SNPs with the following parameters: SNP mode only (-s), for canonical transcripts (-canon), and annotation of loss of function and non-sense decay switched on (-lof) and up/down stream regions set to 2000bp (-ud 2000). The Grantham score for non-synonymous SNPs was determined by the Grantham matrix table [231] using a custom perl script. All non-synonymous SNPs were then classified in four groups according to their Grantham score: conservative for scores below 50, moderately conservative for score between 50 and 100, moderately radical for score between 100 and 150, and radical for score above 150.

#### **F. Axiom genotyping**

A list of SNPs was submitted to Affymetrix for probeset design. For each each SNP, one single probeset was included in the TaBW280K SNP array. Genotyping was conducted on the Affymetrix GeneTitan system according to the procedure described by Affymetrix (Axiom® 2.0 Assay Manual Workflow User Guide Rev3). Allele calling was carried out using a modified version of the Affymetrix proprietary software packages Affymetrix Power Tools (APT) and SNPolisher™ ([http://www.affymetrix.com/estore/partners\\_programs/programs/developer/tools/devnettools.affx](http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/devnettools.affx)) to take into account the specificities of the wheat genome. The HomFLD filter was set to 3.6. As a first step, all the probesets were processed with a mild inbred penalty equal to 4 on all the samples. As a second step, the SNPs failing the QC criteria (“Other” and “NoMinorHom”) were reprocessed using an inbred penalty of 16. Probesets classified as OTVs by SNPolisher were analyzed with OTV\_caller in the two steps. The TaBW280K is available upon request.



## G. Variability estimation and diversity analysis

Variability for each locus was measured using the Polymorphism Index Content (PIC) [161]

$$: PIC = 1 - \sum_i^n p_i^2$$

where  $p_i$  is the frequency of the  $i$ th allele. Genotyping data were used to describe diversity within the wheat panel. A dissimilarity matrix was built using simple matching coefficient between each pair of accessions and the diversity was analyzed by a Ward dendrogram and a Neighbor-Joining Tree. Data analyses were conducted using the DARwin software [267,268].

## H. Genetic mapping

PHR SNPs in the Chinese Spring x Renan F6 population were selected and filtered out to select diploidized probesets ( $\text{HomFLD} > 10$  and  $\text{HomRo} > 0.3$ ). A second filtration was applied to discard markers which significantly ( $P \leq 0.01$ ) deviated from the expected 1:1 ratio in a chi-square test, markers with missing or heterozygous data in parents, markers with more than 15% missing data. SNPs were divided then into 21 different sets corresponding to the 21 chromosomes, based on their CSS-based *in silico* assignment. Genetic maps were then constructed using MSTmap [269] with the following default parameters: population type: RIL6; distance function: Kosambi; cut-off: 0.0000000001; map dist.: 15; map size: 2; missing threshold: 0.20; estimation before clustering: yes; detect bad data: yes; objective function: ML. Once robust framework maps were obtained for each chromosome, a whole-genome map was built using a set of 5,230 selected markers covering all chromosomes and genetic bins. This map was used to place additional markers consisting in unassigned markers from previous chromosome-per-chromosome analyses, as well as markers that were heterozygous for one parent, markers with missing data in parents and that could be manually re-assigned, and distorted markers that were excluded during the second filtration phase.



## Author Contributions

Conceptualization: N.R., C.F., P.S. and E.P.; Methodology: H.R., B.D., M. L., J.D., C.P., P.S. and E.P.; Software: H.R., F.C. and A.D. ; Formal analysis: H.R., B.D., J.N., J.K. and E.P.; Investigation: H.R., B.D., J.D. and J.K.; Resources: IWGSC, M.C.P., D.B. and T.B.C; Writing – Original Draft: H.R., B.D. and E.P.; Writing – Review & Editing: F.C., P.S. and E.P.; Supervision: N.R., P.S. and E.P.; Project administration: T.B.C., P.S. and E.P.; Funding acquisition: T.B.C., C.F. and E.P.

## Acknowledgments

The research leading to these results have received funding from the French Government managed by the Research National Agency (ANR) under the Investment for the Future programme (BreedWheat project ANR-10-BTBR-03), from FranceAgriMer, French Funds to support Plant Breeding (FSOV) and from INRA. Axiom genotyping was conducted on the genotyping platform GENTYANE at INRA Clermont-Ferrand ([gentyane.clermont.inra.fr](http://gentyane.clermont.inra.fr)). Seeds from the INRA core collection were provided by the Small Grain Cereals Biological Resources Centre at INRA Clermont-Ferrand ([http://www6.clermont.inra.fr/umr1095\\_eng/Teams/Research/Biological-Resources-Centre](http://www6.clermont.inra.fr/umr1095_eng/Teams/Research/Biological-Resources-Centre)). Thanks to Anne Boland, Marie-Thérèse Bihoreau and their staff (CEA-IG/CNG) for their technical assistance in DNA quality check and Illumina sequencing.



## Supplemental Information (SI)

**Supplemental Table S1.** List of 3,289,847 intervarietal SNPs with fraction (ISBP or low-copy sequence), chromosomal origin, corresponding IWGSC contig, and context sequence.

**Supplemental Table S2.** The TaBW280K SNP array. List of 280,226 SNPs with probeset ID, SNP ID, fraction (genic or intergenic), chromosomal origin, corresponding IWGSC contig, context sequence, alleles, SNP category (PHR, OTV, MHR, NMH, CRBT and Other), numbers of AA, AB, BB, OTV and missing data, and PIC value.

**Supplemental Table S3.** List of 96 wheat lines genotyped with the TaBW280K SNP array

**Supplemental Table S4.** A 53K loci genetic map of the Chinese Spring x Renan population





## 2. Principales conclusions

Les données de ré-séquençage de huit variétés ont permis le développement de plus de 3.3 millions de SNP distribués dans les portions géniques et les portions non répétées inter-géniques du génome du blé.

De plus, une partie de ces SNP a fait l'objet d'un transfert vers une puce de génotypage haut débit de type Axiom affymetrix® TaBW420k contenant 423 385 SNP dont 280 226 publics (Axiom 280k dans l'article). Afin, de maximiser les chances de transfert des SNP détectés, les critères suivants ont été appliqués : SNP présent dans au moins deux des huit lignées, polymorphe dans les lignées européennes et présent en une seule copie lors de l'alignement avec les séquences survey des chromosomes.

Ensuite, la puce Axiom affymetrix® TaBW420k a permis la caractérisation de 96 accessions de blé tendre au travers d'analyses phylogénétiques en utilisant une analyse basée sur le polymorphisme de présence/absence. Enfin, la puce aura aussi permis la construction d'une carte génétique haute densité qui comprend 53 287 marqueurs sur la population CsRe.

Finalement, cet article fournit à la communauté une large collection de SNP ainsi qu'une puce de génotypage à haut débit et une carte génétique haute densité calculée sur une seule population dont un des parents est la variété internationale utilisée pour le séquençage (Chinese Spring).

Tableau 2 : **Récapitulatif des données de la carte BreedWheat**

A

		A	B	D
1	Génétique	13 201	14 210	4 497
	BioInfo	5 770	6 355	1 540
	Total	18 971	20 565	6 037
2	Génétique	14 420	13 164	6 392
	BioInfo	6 702	7 262	2 466
	Total	21 122	20 426	8 858
3	Génétique	12 200	16 446	5 143
	BioInfo	5 024	7 315	1 408
	Total	17 224	23 761	6 551
4	Génétique	11 750	9 267	2 937
	BioInfo	5 837	4 144	806
	Total	17 587	13 411	3 743
5	Génétique	10 599	13 385	3 699
	BioInfo	5 098	6 424	1 907
	Total	15 697	19 809	5 606
6	Génétique	10 217	14 169	5 369
	BioInfo	4 846	6 134	1 814
	Total	15 063	20 303	7 183
7	Génétique	13 892	12 279	5 348
	BioInfo	6 095	5 891	1 848
	Total	19 987	18 170	7 196

B

Génétique (CsRe + ITMI)	212 584
Informatique	94 686
Total	307 270

(A) Récapitulatif de la position et du nombre de marqueurs positionnés sur les 21 chromosomes du blé tendre. Génétique : marqueurs génétiquement positionnés à l'aide des cartes Chinese Spring X Renan et ITMI. BioInfo : marqueurs positionnés grâce à l'information de séquence du chromosome survey. (B) Bilan total des marqueurs.

### 3. Résultats complémentaires

#### A. La cartographie du projet BreedWheat

Dans le cadre du projet BreedWheat j'ai été amené à réaliser une combinaison de différentes cartes génétique à partir des données de génotypage des puces Affymetrix Axiom TaBW420k (Axiom 280k dans l'article pour les SNP publics). Finalement, j'ai positionné 307°270 marqueurs sur les 21 chromosomes du blé tendre à l'aide de deux populations et d'outils informatiques (Tableau 2). Les deux populations de cartographie sont Chinese Spring X Renan et ITMI. La population Chinese Spring X Renan, développée dans le laboratoire se compose de 1270 lignée sélectionnées en Single seed descent. ITMI est une population de cartographie publique qui résulte du croisement entre W7984 (blé synthétique issue du croisement entre *Ae. tauschii* x *Altar durum*) x Opata 85 (blé tendre mexicain) [236]. Ce croisement a été conçu pour maximiser le polymorphisme (notamment RFLP) sur le génome D.

Les 307°270 marqueurs se répartissent en : 94°686 positionnés par bioinformatique, 29°265 issus de données de génotypage de la population W7984 x Opata (ITMI) et 183°319 issus de la population Chinese Spring X Renan utilisée comme référence pour l'établissement de la carte finale. Le total des marqueurs ordonnés à l'aide de leur position génétique était de 212°584 dont seulement 159 marqueurs présentant des positions différentes entre les populations ITMI et CsRe ce qui suggère des réarrangements chromosomiques entre les différents parents des deux populations.

Tableau 3 : **Récapitulatif des données de la carte PolyHigh**

		A	B	D	Total Marqueur	Total cM	Total Bin
1	Marqueur	6813	7147	2459	16419	1093	438.4
	cM	146.8	133.2	158.4			
	Bin	459	398	236			
2	Marqueur	7494	5621	2985	16100	905	389.0
	cM	107.1	131.3	150.6			
	Bin	316	335	254			
3	Marqueur	6159	7838	2632	16629	1166	520.3
	cM	156.4	164.0	200.0			
	Bin	371	449	346			
4	Marqueur	5916	4219	1591	11726	914	443.6
	cM	181.4	113.1	149.1			
	Bin	316	276	322			
5	Marqueur	5465	6392	1860	13717	1177	591.9
	cM	193.4	178.1	220.4			
	Bin	451	390	336			
6	Marqueur	4958	7578	2723	15259	838	424.4
	cM	127.2	112.6	184.6			
	Bin	253	315	270			
7	Marqueur	7073	5461	2623	15157	1129	501.1
	cM	176.3	126.4	198.4			
	Bin	456	336	337			
	Total Marqueur	43878	44256	16873	105007		
	Total cM	1088.5	958.781	1261.346			3308.665
	Total Bin	2622.0	2499	2101		7222	

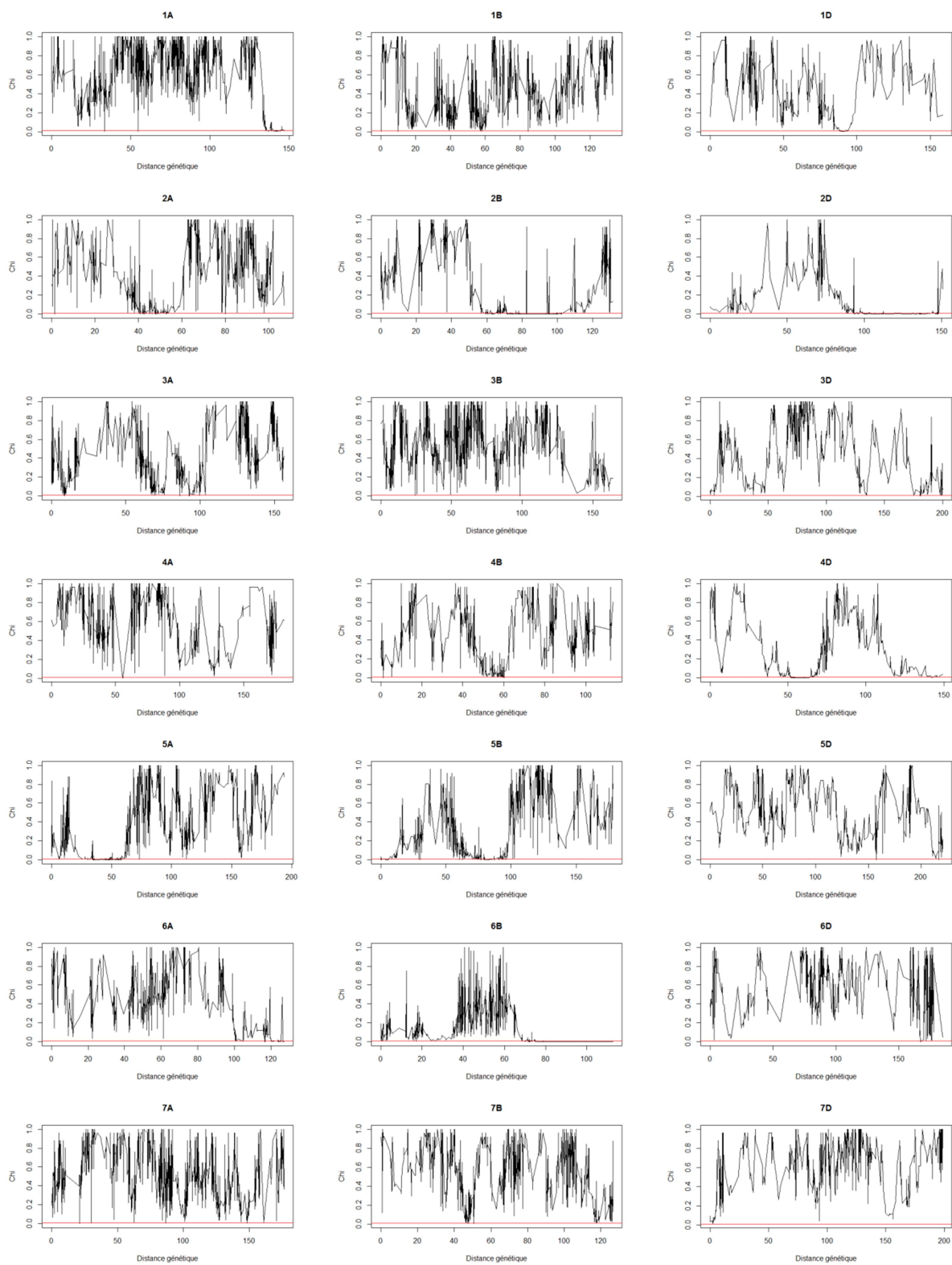
Récapitulatif du nombre de marqueurs génétiques cartographiés et de la distance génétique calculée pour chacun des 21 chromosomes du blé tendre. Le nombre de bins génétiques (nombre de groupe de marqueurs qui co-ségrègent), ainsi que le nombre total de marqueurs et la distance génétique totale sont indiqués. De plus, sont présentées les valeurs totales par génome et par groupe de chromosomes homéologues.

## B. Chinese SpringX Renan : PolyHigh

L'étude de la recombinaison et les analyses que j'ai réalisées ont utilisé une version allégée et robuste de la cartographie génétique du génome entier composée de 105°007 marqueurs PolyHighResolution (Tableau 3). Ces marqueurs sont classés comme « PolyHighResolution » en fonction de la facilité d'analyse des données de la puce. Cette catégorie regroupe les marqueurs qui franchissent tous les seuils statistiques pour assurer une bonne détermination du génotype. Néanmoins, ces analyses restent occultes et dépendantes de la société Affymetrix.

La carte générée à l'aide des marqueurs PolyHighResolution est appelée PolyHigh. Elle couvre les 21 chromosomes du blé tendre, représente une distance de 3308 cM répartis dans 7222 bins génétiques.

Lors de la réalisation de la cartographie, 5 chromosomes ont présenté des biais de ségrégation importants : les chromosomes 2B (59% des marqueurs du chromosome entre les positions 60 à 100 cM), 2D (45% ; 100-147 cM), 4D (33% ; 42-63 cM) et 5A (21% ; 35-60 cM) en faveur des allèles de Renan et le chromosome 6B (12% ; 70-112 cM) en faveur de Chinese Spring. Le nombre maximal de marqueurs placés a été trouvé pour le chromosome 3B avec 7838, tandis que le nombre minimal de marqueurs est de 1591 pour le chromosome 4D. En moyenne le nombre de marqueurs génétiques par chromosome est de 6268 pour le génome A, 6322 pour le génome B et de 2410 pour le génome D. Le nombre maximal de marqueurs génétiques est observé pour les groupes d'homéologie 1 et 3 avec 16419 et 16629 marqueurs respectivement. Le nombre le plus faible est observé pour le groupe 4 avec 11726 marqueurs.



**Figure 33 : Distribution des valeurs de  $\chi^2$  le long des chromosomes**

Représentation des valeurs de  $\chi^2$  pour chaque marqueur le long de la carte génétique des 21 chromosomes du blé tendre.

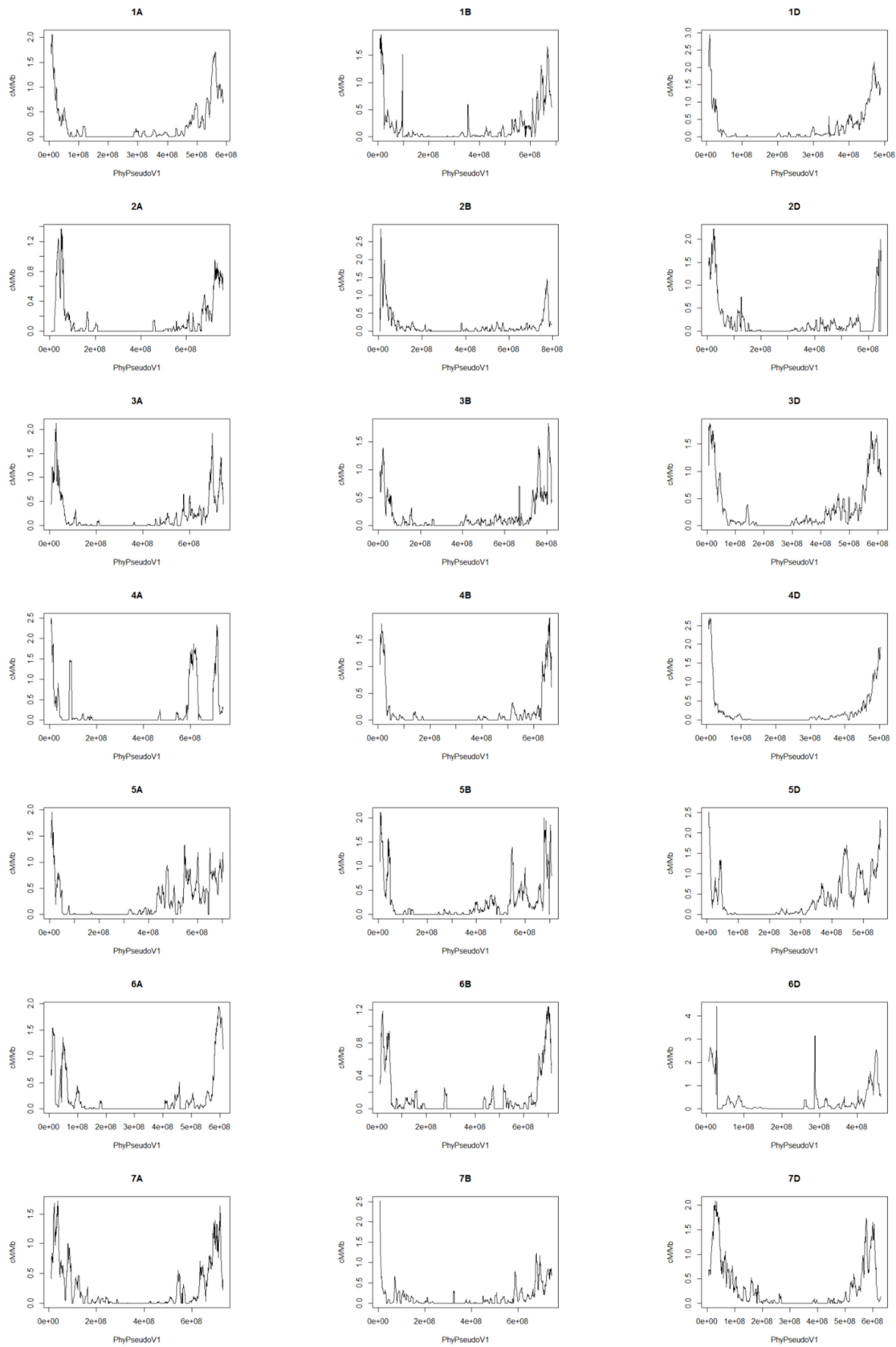
Concernant la taille génétique par chromosome, celle-ci est en moyenne de 155.5 cM pour le génome A, 137 cM pour le génome B et de 180 cM pour le génome D. Le maximum de distance génétique est observé pour les groupes d'homéologie 5 et 3 avec respectivement 592 et 520 cM. Le minimum est observé pour les groupes 2 et 6 avec 389 et 424 cM respectivement.

En termes de bins génétiques, les moyennes des différents génomes sont de 374 pour le génome A, de 357 pour le génome B et de 300 pour le génome D. Il apparaît aussi que les groupes 5 et 3 possèdent le plus de bins de génétiques avec respectivement 1177 et 1166 bins génétiques. Le groupe qui possède le moins de bins génétiques est le groupe 6 avec 838 bins.

### **C. Biais de ségrégation**

A la lumière des différentes distorsions de ségrégation présentes sur 5 chromosomes, l'ensemble des valeurs de  $\chi^2$  a été visualisé chromosome par chromosome en utilisant la distance génétique de chaque marqueur comme abscisse. Ainsi, plusieurs chromosomes présentent de longues étendues en distorsion de ségrégation non significative. Certaines de ces zones semblent restreintes aux bras des chromosomes comme pour les chromosomes 2B, 2D, 5A, 6A, 3B alors que d'autres zones semblent englober le centromère comme pour les chromosomes 5B ou 4D par exemple. (Figure 33)





**Figure 34 : Taux de recombinaison le long des 21 chromosomes du blé tendre**

Courbe du taux de recombinaison en cM/Mb le long des cartes physiques NRGene des 21 chromosomes du blé tendre (fenêtre glissante de 10 Mb avec un pas de 1 Mb).

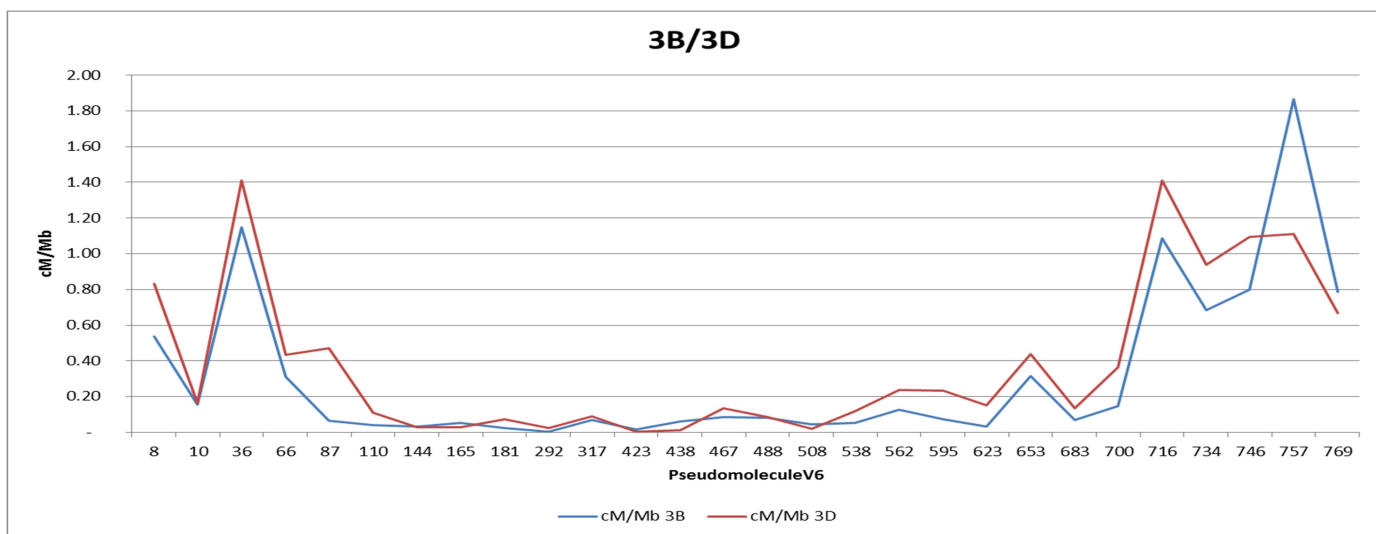
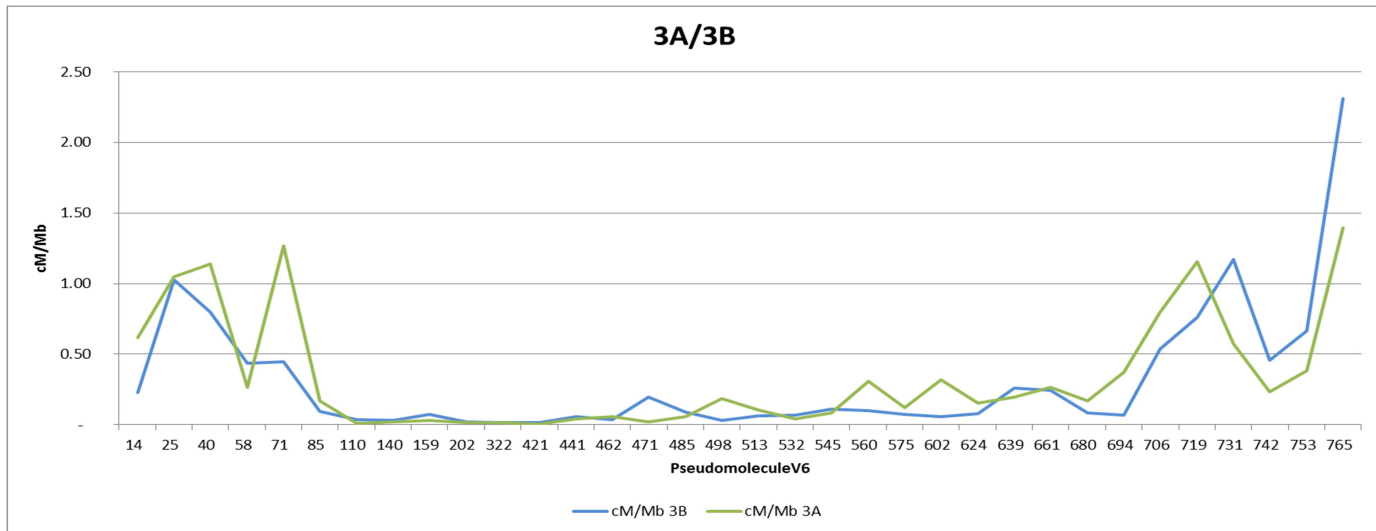
#### **D. Analyse de la recombinaison génome entier**

L'accès à une séquence de qualité au travers des nouveaux assemblages NRGene m'a permis de réaliser l'analyse de la distance physique par rapport à la distance génétique à l'échelle du génome entier à l'aide de mes cartes génétiques (Figure 34). Il apparaît ainsi que le partitionnement de la recombinaison développé sur le chromosome 3B semble être une généralité de tous les chromosomes du blé tendre. J'ai pu comparer de manière plus précise le groupe d'homéologie 3 en utilisant la séquence du chromosome 3B comme séquence de référence (Figure 35). Le résultat montrait un profil de recombinaison partitionné avec les mêmes régions qui semblent recombiner chez les 3 homéologues mais à des intensités différentes.

#### **4. Discussion sur les résultats complémentaires**

Les analyses du génome entier du blé dans le cadre du projet BreedWheat m'ont permis de réaliser une carte génétique de référence du blé tendre. Cette carte de référence réalisée pour la population Chinese Spring X Renan a montré plusieurs biais de ségrégation.

Les biais ou distorsions de ségrégation sont un écart à la distribution normale attendue de  $\frac{1}{2}$  et  $\frac{1}{2}$  dans le cas d'une population de lignées recombinantes en génération F6 à deux parents (3.125% d'hétérozygotie résiduelle). Cet écart témoigne peut-être d'une sélection d'un des allèles parentaux. Cette sélection peut-être une sélection humaine ou biologique. Dans le cas de la population Chinese Spring X Renan, la sélection humaine a été réduite au maximum. Néanmoins compte tenu du mode de descendance choisie, un possible biais pourrait exister lors de la génération 3 car plusieurs plantes F2 n'ont pas présenté de graines ou étaient décédées. Cette possible sélection ne suffit pas expliquer la présence de plusieurs biais de ségrégation. Par conséquent des causes géniques doivent être envisagées.



**Figure 35 : Taux de recombinaison en cM/Mb le long de la séquence de la pseudomolécule V6 pour le groupe d'homéologie 3**

Représentation du taux de recombinaison le long de la séquence du chromosome 3B pour les chromosomes 3A/3B/3D. Les chromosomes 3A et 3D ont été ancrés sur le chromosome 3B en utilisant les séquences contextes des SNPs.

Par ailleurs, la distorsion du chromosome 2B a déjà été observée dans différents fonds génétiques [270–272] mais elle ne peut pas être attribuée à des facteurs androgéniques comme suggéré dans [272] car le type de descendance ne le permet pas. Néanmoins, un gène impliqué dans la méthylation de l'ADN (Met1) est présent sur le chromosome 2B [273] et une variation allélique de ce gène pourrait peut-être entraîner cette distorsion. L'analyse des différentes copies du gène Met1 a montré que celles du groupe 2 contribueraient de manière majoritaire à son expression et que la copie du chromosome 2B semble être la plus exprimée par rapport aux copies 2A et 2D. Les copies du groupe 2 semblent de plus jouer un rôle dans l'extinction des autres copies présentes sur les groupes 5 et 7 [273].

Ainsi, il existe une très forte pression de sélection sur le chromosome 2B qui pourrait conduire à un biais de ségrégation dans le cas d'une hybridation entre deux blés tendres. La présence d'un gène de type pollen killer (*Ki*) est également relatée sur le chromosome 6B [274]. Dans le cas du chromosome 4D, un biais de ségrégation a été trouvé par [275] sans que les auteurs avancent une quelconque explication. Néanmoins, des gènes de stérilité mâle se trouvent sur ce chromosome et sont très liés génétiquement à un gène de nanisme [276,277]. Finalement, la cause des biais de ségrégation est difficile à déterminer mais leur apparition témoigne de la présence de pression de sélection. Dans le cas de la population Chinese Spring X Renan, les observations sont effectuées sur la génération 6. Il serait donc peut-être intéressant d'observer le comportement méiotique des générations F2 afin de voir si certains gamètes présentent des structures anormales qui impliquent les chromosomes qui montrent des biais de ségrégation.



La taille de la carte génétique de la population Chinese Spring X Renan se situe dans la moyenne basse de la taille du génome blé tendre. Cette information semble montrer soit l'absence de marqueurs dans certaines parties du génome soit une moindre recombinaison par rapport à certaines populations sur certains chromosomes. C'est le cas notamment du groupe d'homéologie 2 qui présente des distances génétiques réduites. Ce niveau très faible de recombinaison peut être le fait d'introgessions provenant d'*Aegilops ventricosa*. En effet, Renan, variété déposée au catalogue en 1989 et développée par l'INRA, est issue d'un croisement complexe et possède VPM1 comme parent. Celui-ci présentait des introgessions d'*Aegilops ventricosa* pour des locus de résistance qui se retrouvent dans le bras court du chromosome 2A [278].

De plus, la taille des cartes génétiques montre des différences d'intensité de recombinaison entre les génomes. En effet, le génome D recombine plus que le génome A qui recombine d'avantage que le génome B dans la carte PolyHigh. Cette information semble être en accord avec le taux de polymorphisme des génomes [143] qui est maximum dans le génome B et très limité dans le génome D. La divergence entre séquences est connue pour limiter la recombinaison [279,280].

Finalement, les résultats de l'analyse de la recombinaison à l'échelle du génome entier ont montré que la recombinaison semble avoir lieu dans les mêmes zones physiques (parties distales) chez tous les chromosomes. Par ailleurs, l'étude à l'échelle des chromosomes du groupe 3 semble montrer que la recombinaison se situe dans les mêmes régions mais à des intensités différentes ce qui suggère qu'elle a lieu dans les régions les mieux conservées entre chromosomes homéologues à savoir, les parties codantes.



## 5. Conclusion

La réalisation de ces cartes génétiques m'a permis de participer au projet BreedWheat et de fournir aux sélectionneurs un outil génomique pour positionner/identifier des gènes d'intérêt majeur. De plus, ces cartes sont utiles pour l'ancrage des cartes physiques et de la séquence du génome du blé tendre et ont notamment été utilisées pour vérifier/aider à l'assemblage des séquençages NRGene.

Finalement, ces travaux de cartographie ont été d'un grand intérêt pour l'étude de la recombinaison à l'échelle du génome entier et ainsi montrer que le chromosome 3B pouvait être un chromosome modèle. Grâce à la réalisation de la cartographie PolyHigh et des données physiques des contigs, plus de 500 CO ont été localisés avec précision sur l'ensemble du génome et ont pu être utilisés directement dans la suite du projet de thèse.





## **IV. Partie III : Etude des paramètres qui impactent la distribution de la recombinaison sur le chromosome 3B et l'ensemble du génome**

---

### **1. Introduction**

Les résultats concernant la distribution de la recombinaison le long de la pseudo-molécule du chromosome 3B ont permis d'identifier 22 régions, possédant un taux de recombinaison supérieur à 10 fois la moyenne du chromosome, pouvant être considérées comme des points chauds de recombinaison [121]. Ces points ont servi de base à la sélection des régions à analyser plus finement pour l'étude des facteurs impactant la distribution des crossing-overs. Ce travail fait l'objet du dernier article accepté pour publication dans la revue Genetics.



# High-resolution mapping of crossover events in the hexaploid wheat genome.

---

## 1. Authors

Benoit Darrier\*, H el ene Rimbert\*, Fran ois Balfourier\*, Lise Pingault\*, Ambre-Aurore Josselin\*, Bertrand Servin†, Julien Navarro\*, Fr ed eric Choulet\*, Etienne Paux\*, Pierre Sourdille\*

## 2. Affiliations

\*GDEC, INRA, Universit  Blaise Pascal, 63039, Clermont-Ferrand, France.

†GenPhySE, INRA, Universit  de Toulouse, 31320, Castanet-Tolosan, France

Running Title: Recombination analysis in wheat

Key words: Recombination; Bread wheat; Sequence features; Historical-mapping; Transposable-element-related hotspot motif

Corresponding author: Pierre SOURDILLE

*Genetic, Diversity and Eco-physiology of Cereals, INRA, Site de Crou l, 5 Chemin de Beaulieu, 63039, Clermont-Ferrand, France*

Phone: 33 4 73 62 44 37

Fax: 33 4 73 62 44 53

E-mail: [pierre.sourdille@clermont.inra.fr](mailto:pierre.sourdille@clermont.inra.fr)



### 3. Abstract

Meiosis is an essential process required to create haploid sexual cells. During meiosis, recombination occurs and allows creating new allele associations by reciprocal exchange of DNA. In bread wheat (*Triticum aestivum* L.), Crossovers are almost limited to subtelomeric regions of chromosomes resulting in a substantial loss of breeding efficiency in the proximal regions though these latter carry ~60-70% of the genes. Identifying sequence and/or chromosome features impacting recombination occurrence is thus relevant to improve and drive recombination in a near future. With the recent release of a reference sequence of chromosome 3B and of the draft assemblies of the 20 other wheat chromosomes, we were able to perform a fine-scale mapping analysis of Crossovers. We genotyped a large segregating population of 1270 F6 lines with 774 SNPs highlighting 74 small intervals (<26 kb) on chromosome 3B that precisely delineate 252 crossover events. Additional genotyping of two collections of 180 varieties representative of the Asian and European genetic pools with the same SNPs, revealed a common location for ancestral recombination events (predicted through Linkage Disequilibrium analyses) and the previously-identified hotspots. Surrounding-crossover-sequence analysis showed that Crossovers occur more frequently in gene-rich regions and around genes expressed during meiosis. We also observed differences in transposable element content and we identified an original associated motif (carried by *TIR-Mariner transposon*) that might drive recombination. This motif shares high degree of similarity with the one associated with recombination hotspots in human and carried by *THE1A/B* retrotransposon, suggesting a common ancestral mechanism for recombination control in eukaryotes.



#### 4. Introduction

Meiotic recombination is a process that allows reshuffling of diversity by reciprocal exchange of DNA called crossover (CO). This phenomenon is conserved in most eukaryotes (for a review see [16]) and follows double strand break (DSB) of DNA generated by the topoisomerase *SPO11* complex. Recombination is highly regulated and at least one CO per chromosome is absolutely required to ensure a correct segregation of chromosomes during meiosis [16]. CO location can be determined either through their cytological signature (chiasmata) or through the parental allele switch in progenies (for a review see [48]). This latter can be detected by genotyping the progenies or from the resulting Linkage Disequilibrium (LD) patterns in populations [281]. In all eukaryotes, the distribution of COs is not homogeneous along the chromosomes with a drastic reduction of recombination at the centromeres [16,63,66,109,110,121,282,283]. Recombination is thus restricted to a few percent of the chromosomes representing no more than one fourth of the genome in *S. cerevisiae*, *Arabidopsis*, wheat and humans [16,77,109,143]. For cultivated crops, this implies a decrease of the selection power in regions showing low CO rates [284].

Previous studies aiming at deciphering recombination in *Arabidopsis* and wheat evaluated 540-23,000 bp intervals and revealed the presence of small regions termed “hotspots” [16,63,71,76,106] where recombination ranges from 8.82 to 75.52 cM/Mb compared to average chromosome values of 4.8 and 0.20 respectively [63,78,79]. However, the number of DSBs is at least 10-50 fold greater than the number of COs which rarely exceeds three per bivalent chromosome per meiosis [16]. This paucity of crossovers per meiosis with regards to the number of DSBs suggests the establishment of clear driving-factors of recombination in plants as well as consistent relationships between recombination rates and sequence features.





In *Arabidopsis* [77], historical mapping analysis shows CO-frequency increases towards transcription start (TSSs) and termination (TTSs) sites of the genes with a stronger effect for TSSs. Thereby hotspots were localized in promoters and terminators of genes and four were confirmed by allele-specific amplification from gamete DNA (pollen typing) [77–79]. In maize, fine scale studies at the *bronze* locus revealed that recombination occurs preferentially in or around genes and that retrotransposons are probably inert with regard to recombination because of methylation and the likely associated condensed chromatin [73,75]. In sorghum, a similar correlation was observed with 97-98% of COs occurring in euchromatin regions [141]. However, many gene-rich regions show little or no recombination in plants as well [79].

Contradictory results are also observed regarding the impact of transposable element (TE) content on recombination. Repeated sequences and recombination rate are correlated positively in *C. elegans* [285] and negatively in *D. melanogaster* genome [286]. Even in the same species, in maize, a reduction of recombination was observed in regions containing transposable elements (TEs) [75,284] while insertion sites of *Mu* transposon concentrate in the same regions as meiotic recombination [88].

Transposon abundance, diversity and activity are highly variable among species and they developed diverse strategies to thrive in host such as preferential insertion close to genes [91,287]. TEs contribute to chromosome shape and function (for a review see [92]). Recently, a study in mouse revealed that DNA methylation can prevent transposons from adopting chromatin characteristics amenable to meiotic recombination [93]. In *Arabidopsis*, epigenetic marks can reduce recombination by increasing nucleosome density [288]. This underlies the importance of DNA conformation and the likely role of epigenetic marks on recombination.



Mouse and Human use the same molecular strategies regarding the regulation of recombination. DSBs occur in regions enriched in H3K4me3 epigenetic mark which is known to be controlled by *PRDM9*. Initially in Human, CCTCCCT motif was found overrepresented at hotspots in the *THE1A/B* retrotransposon which is primate-specific [86,87,89,289]. Additional data on multiple families' backgrounds in Human revealed the presence of a common 13-bp degenerated motif CCNCCNTNNCCNC. In plants, no homolog to *PRDM9* subfamily was found [94] and no particular motif was found to explain the presence of hotspots. However, most H3K4me3-methylase encoding genes are expressed in meiotic cells in *Arabidopsis* and rice [94] and this mark is associated with recombinogenic regions in *Arabidopsis* and barley [77,95,98]. This may even be a general eukaryotic regulatory mechanism and may have originated in the ancestor of eukaryotes [16,94].

Frequency of meiotic recombination is highly biased toward the ends of chromosomes in large genomes like in bread wheat (*Triticum aestivum* L.;  $2n=6x=42$ ; 17 Gb) [112–114,158,251]. However, this does not rely on the distal position since when the chromosome arms are inverted (telomeres placed at the centromeres and *vice versa*), COs still occur in the same region whatever its location [70], suggesting a DNA factor or chromatin organization promoting recombination in distal regions. Despite complexity of the bread wheat genome, a hotspot was already reported in the promoter of *TaHGA3* gene on chromosome 3B [76]. Recently, the reference sequence of chromosome 3B and the draft assemblies of the 20 other chromosomes were published [121,143], offering the opportunity to precisely characterize recombination hotspots at the whole-genome scale.



In this work, we used these new resources to get a genome wide overview of recombination and to get better insights into the factors that drive crossovers in wheat. First, we fine-mapped COs on chromosome 3B using a large segregating population. We showed that these current recombination hotspots are conserved with ancestral recombination breakpoints determined through historical mapping (termed Rho-Map herein; [290]) using two collections (European and Asian) representing two different genetic pools of bread wheat. Our analysis of >250 COs located on chromosome 3B revealed that recombination mainly occurs nearby or within genes which are mostly expressed during meiosis. Furthermore, TE composition is different between recombinogenic or non-recombinogenic intervals and TE-related DNA motifs are associated with high recombination rates suggesting a possible ancestral mechanism for recombination in eukaryotes.

## **5. Materials and methods**

### **A. Plant material**

CsRe population was derived from a cross between Chinese Spring (international reference) and Renan (a French elite cultivar). The progeny was developed through single seed descent until F6. Two sets of 276 and 356 lines were randomly selected to genotype 5778 (Axiom Array®) and 1280 (KASPar®) SNPs respectively while the whole population (1270 individuals) was genotyped with only 96 SNPs to roughly identify the recombinants along the pseudomolecule. Concerning the two contrasted collections made of 90 Asian accessions and 90 European ones (Table S3), those accessions were chosen within the Asian wheat gene pool and the North-western European wheat gene pool, which are genetic pools of origin for Chinese Spring and Renan cultivars, respectively [233]. The two sets of 90 accessions were selected in order to represent the diversity within their considered genetic pools.



## B. Genotyping data and genetic mapping

SNPs were developed using the Chinese Spring survey sequences from [143] as reference to align the reads from Renan and make the SNP calling. We identified 192,584 SNPs on chromosome 3B that were polymorphic between Chinese Spring and Renan. SNPs were assigned *in silico* to the 3B pseudomolecule and unique loci with 100% of overlap and 100% identity were selected leading to 136,458 predicted SNPs. Finally, 22,680 supplementary SNPs derived from other projects (Axiom Array®) were added to improve our capacity of SNP development. In addition, 168,211 Simple Sequence Repeat (SSRs) were detected on the pseudomolecule using an in-house pearl script.

In total 774 SNPs and 61 SSRs (Table S4) were used on the CsRe population among which 466 SNPs were derived from [291], 214 from [121], 88 transferred from Axiom® 420k array [291] and 6 from previous studies of recombination [76]. 540 SNPs and 9 SSRs were retained to refine CO localization. For the two Asian and European collections, only 96 SNPs were used. Genotyping data using SNPs were acquired using KASPar technology. Triplex of primers were designed manually according to manufacturer instructions using informatic prediction and context sequences. Two techniques were used to process to genotyping: Fluidigm® technology (microfluidics) & Roche LightCycler480®. SSR marker polymorphisms were analyzed on 1% agarose gel. Whole genome genotyping data were obtained through Axiom® array as described in [291]. Genetic mapping (105,183 SNPs mapped on the 21 chromosomes) was achieved according to [291].





### C. CO detection

A CO was defined on each line as a switch of parental alleles between two markers of known physical position. According to our strategy we used scaffold information resulting in mapping/detection of markers along chromosome 3B pseudomolecule [121] to locate and refine CO position between 2 markers present on the same scaffold. These strategies allow access to a confident physical position at scaffold scale. To certify quality and avoid chimeric scaffolds, genetic map was realized before refinement of CO position using whole population data. Consistency of genotyping data was verified using graphical genotype to manually order markers according to the genetic rules and the physical position. These approaches revealed a few double allele swaps. However they were explained by only one SNP and the use of an F6 progeny did not allow us to discriminate between an NCO and two close CO events. These events were thus removed from our analysis and only one-allele swap supported by neighboring genotyping data contributed to our CO number.

For the whole-genome analysis, we selected all contigs from IWGSC survey sequence which presented an increase of genetic distance between two markers. All genotyping information of these contigs and consistency within global chromosome genetic map were manually verified with graphical genotype to exclude chimeric contigs and putative NCO. We thus confirmed presence of COs supported by neighboring genotyping information and we used physical information available on IWGSC survey assembly [143].



#### **D. Rho-map**

Rho-map was computed using PHASE 2.2 [281,292] and genotyping data derived from the two Asian and European populations. Presence/absence polymorphism was encoded using available multi-allelic function while other SNPs were encoded as bi-allelic markers as described in the PHASE user manual. We used the parameter MR=0 which is the general model for recombination of PHASE 2.2 [281,292]. The MCMC chains were ran for 10,000 iterations (-X100 option). From the software output, we extracted in each interval the posterior distribution of lambda. We used the median of this posterior distribution as an estimate of the interval specific recombination intensity.

#### **E. Sequence information and analyses**

Sequence information from IWGSC chromosome survey contigs (CSS) [143] and the pseudomolecule from chromosome 3B [121] were used for detection of correlation between sequence features and recombination as well as for motif discovery. For chromosome 3B, we used sequence annotation for genes and transposable element described in [121]. For promoter and terminator regions, we used available data when it was already define or we considered 2 kb upstream or downstream after UTR regions. Concerning IWGCS CSS [143], we used The Munich Information Center for Protein Sequences (MIPS) annotation as well as new transcript regions identified by RNA-seq mapping (see below). Transposable elements associated to IWGSC CSS contigs were identified using Repeat masker and ClariTE [123]. We scanned Rec and NoRec intervals for matches with probabilistic matrix provided by MEME Suite 4.10.2 (see below). Sequence analysis was performed on unix system using BEDtools, custom command line/script and R [293] bioconductor [294] suite with genomeIntervals package (v2.17.0 and v2.25.0) [265].

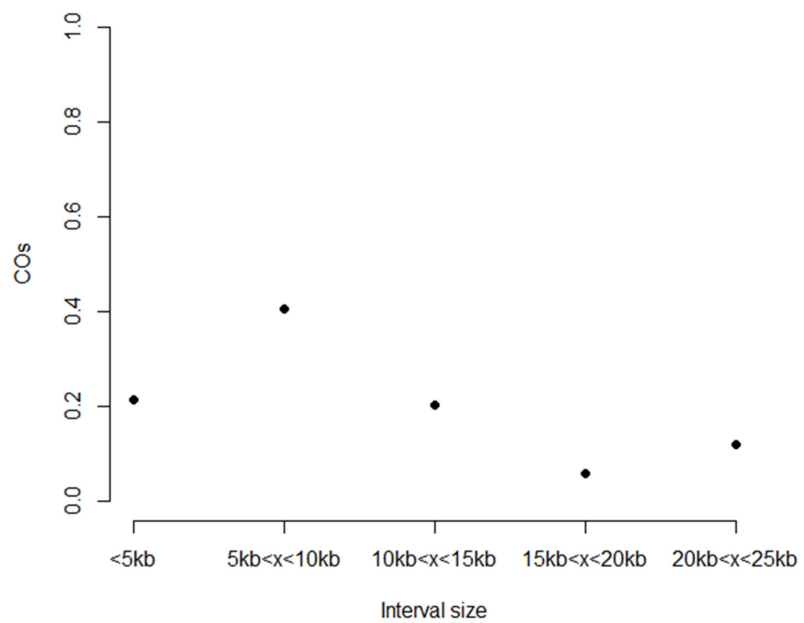


For motif discovery, we used MEME (version 4.10.2) [295] with the following parameters: motifs with nucleotide size ranging between 3 and 20 nucleotides; use first-order Markov model to take into account both nucleotide and dinucleotide (to capture the observed frequency of dinucleotide repeats across the genome) composition; search for motifs on both DNA strands. MEME was also used to generate sequence LOGOS for each discovered motif. Probabilistic matrix computed by MEME & scan\_for\_matches was used to scan DNA sequences to search motif on the basis of the probabilistic matrix.

#### **F. RNA-Seq Mapping**

A total of 656,290,406 RNA-Seq reads from four meiotic stages (Latent/Leptotene; Zygotene/Pachytene; Diplotene/Diakinesis; Metaphase I (data publicly available: <http://wheat-urgi.versailles.inra.fr/Seq-Repository/Expression>) with two replicates were mapped on 86,710 CSS contigs from IWGSC [143] used as a reference with Tophat2 (v2.0.13) with 0 mismatches (-m 0 -N 0 options). Mapped-reads are then filtered using SAMTools and only those mapped with a minimum mapping quality of 30 were kept. Duplicates were removed using SAMToolsrmdup. Mapping data were then used with Cufflinks (v2.2.1) to build GTF files of exons for each condition and each replicates with default parameters. Each GTF file was combined with the gene models from MIPS using cuffmerge (Cufflinks package) with default parameters. Quantification of transcripts was done using cuffquant (Cufflinks package) with 'rescue method' (--multi-read-correct option) and with the output from cuffmerge as transcript reference for each of the BAM files. Finally, cuffnorm was run with default parameters with the GTF file from cuffmerge as reference annotation and with the CXB files from cuffquant. Chromosome 3B meiosis RNA-seq data were established according to [124].

Figure 1



**Fig 1.** Percentage of COs according to interval size in the Rec dataset.

## 6. Results

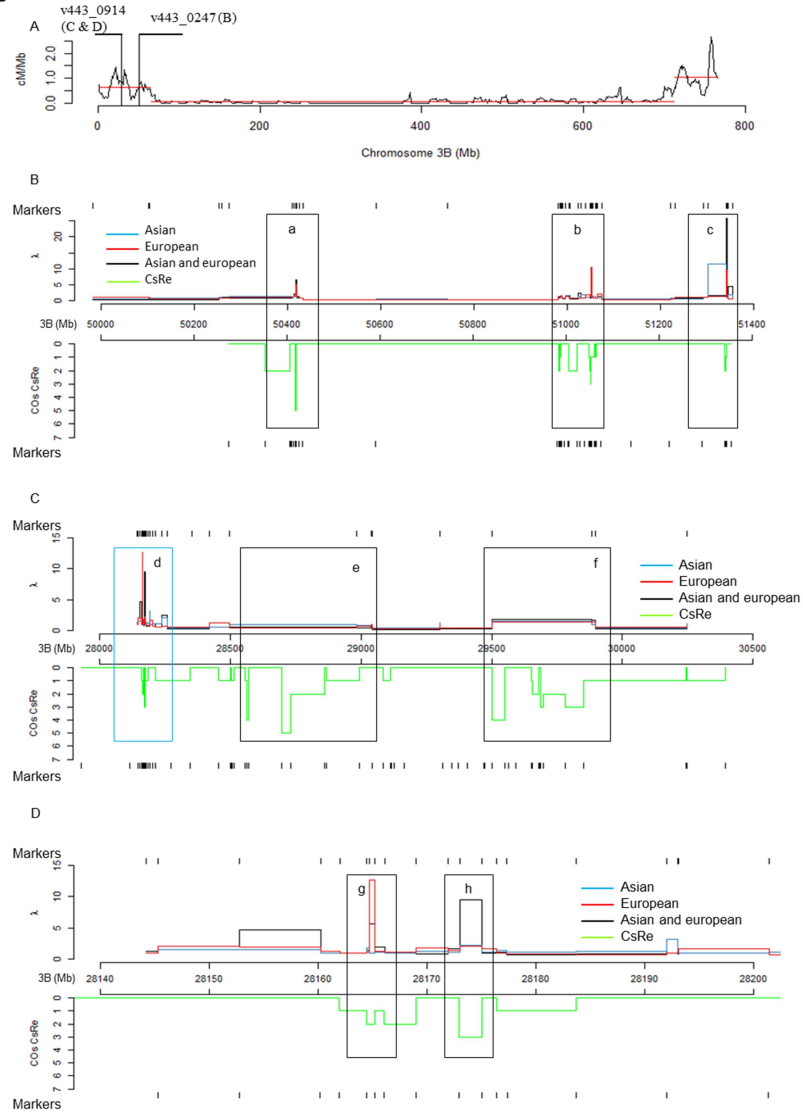
### A. Fine-scale distribution of current recombination on chromosome 3B

We genetically mapped 96 SNP markers from chromosome 3B on 1270 F6 lines derived through single seed descent (SSD) from the cross between Chinese Spring and Renan cultivars (CsRe). In total, we identified 3031 CO events representing 2.38 COs per individual on average for this chromosome. We confirmed the partitioning of recombination with 82% of COs located in the distal ends of chromosome 3B representing 19% of the chromosome length [121]. Using the SNPs available for chromosome 3B [291] on the whole CsRe population, we located 891 COs on 73 scaffolds from chromosome-3B pseudomolecule. These COs were framed by two markers carried by the same scaffold to avoid fallacy physical distance caused by ignorance of gap size or error in ordering of scaffolds. The 73 scaffolds followed the same partitioning as recombination with 72% mapped in the two distal regions and an equal contribution of short/long arms (25 vs 24). The remaining 24 scaffolds were localized in pericentromeric/centromeric regions. (Fig S1). The size of the 73 scaffolds ranged from 101 kb to 2812<sup>o</sup>kb (Table S1).

By further genotyping only lines which presented COs with additional SNPs on 26 among the 73 scaffolds, we narrowed down the resolution of CO location to 74 intervals of less than 26 kb (further called "Rec" to refer to the recombinant dataset) containing 252 COs. The remaining 639 COs could not be accurately mapped in less than 26 kb because: (1) the number of crossovers was found too low in too large features to be selected for additional mapping; (2) the given region lacked polymorphism.



Figure 2



**Fig 2.** Comparative analysis between ancestral and current recombination. Rho-map is computed for: top: Asian (blue) and European (red) genetic pools as well as for the compilation of both (black) using lambda deviation of 0.75 quantile; bottom: segregating population (CsRe; green) using the number of COs observed. Physical location (in kbp) along pseudomolecule is shown in abscises. (A): distribution of recombination rate along the pseudomolecule of chromosome 3B (extracted from [6]). Position of the two scaffolds is indicated; (B): scaffold v443\_0247; (C): scaffold v443\_0914; (D): zoom in on specific area where high resolution mapping is available on scaffold v443\_0914 for both Rho-maps. Breakpoints (a,b,c,d,e,f,g,h) are delimited by black boxes.

Rec intervals ranged from 108 bps to 25,848 bps (average: 9,344°bps). They contained up to 17 COs while 41% carried a single CO. In total, 62% of COs located in intervals lower than 10 kb (Fig 1). Among the 74 Rec intervals, 69 intervals (92%) were found in subtelomeric regions while only five are part of the central region (Fig S2).

Our results confirmed that recombination at the chromosome scale is not homogeneously distributed along the chromosome 3B of bread wheat and mainly occurs (82%) in restricted distal regions (19%).

### **B. Fitting ancestral and current recombination**

Ancestral recombination was estimated by haplotype inference on two populations, each of 90 landraces representing Asian and European genetic pools to fit best with the CsRe bi-parental population (see M&Ms). A recombination background rate was estimated as well as the factor ( $\lambda$ ) by which the recombination exceeds the background rate [281,292]. On the basis of the 74 intervals studied on CsRe, the two largest scaffolds were selected to perform ancestral recombination analysis. The two regions, present on scaffolds v443\_0247 and v443\_0914 (Fig 2) represent 1.2°Mb and 2.5°Mb respectively and are located distally on the short arm (Fig 2A). The mean density of markers in these two scaffolds in the CsRe population was of 1 SNP/28 and 39 Kb respectively. Additional SNPs, derived from a complementing analysis [291], were genotyped in the landraces to achieve a better precision in the estimation of ancestral recombination in these two scaffolds.

For v443\_0247 (Fig 2B), three intervals (a,b,c) exhibiting recombination were identified in CsRe. We observed breakpoints (BPs) overlapping the same three regions in population data. According to the Rho-Map analysis, recombination intensity ( $\lambda$ ) was higher in the European pool for breakpoints (a) and (b) while it was high for both pools for breakpoint (c) resulting in an even stronger deviation from the background rate when pools were analyzed together.



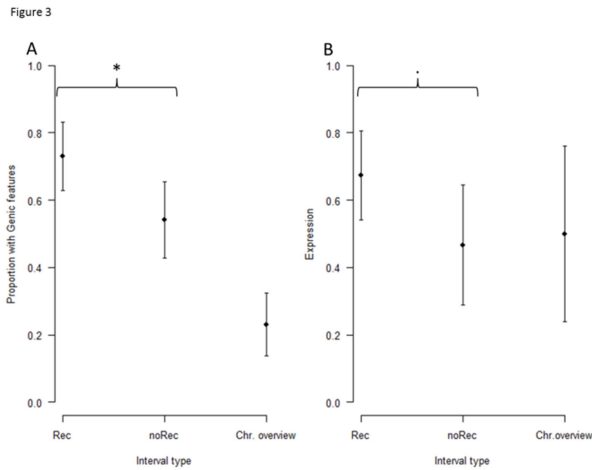
Similarly, for v443\_0914 (Fig 2C) three recombinogenic regions (d,e,f) were identified on the basis of CsRe population and two of them exhibited high  $\lambda$  deviations while no  $\lambda$  deviation from the background rate was observed for the largest interval (e; 484°kb). Interestingly, breakpoint d was highly saturated with SNPs (Fig 2Ca), allowing to reach a very high resolution revealing two historically recombinogenic regions already detected in CsRe (Fig 2D). These two regions (g,h) present strong  $\lambda$  deviation from the background rate similarly to CsRe. However, high  $\lambda$  value was observed only for the European pool for the first one (g) while for the second one (h) elevated  $\lambda$  value was only revealed when both pools were analyzed together.

Although some recombinogenic regions may be unstable along the evolution, ancestral and current recombination patterns are mainly conserved. This suggests that sequence features impacting recombination exist in wheat as demonstrated in *Arabidopsis* [63,284].

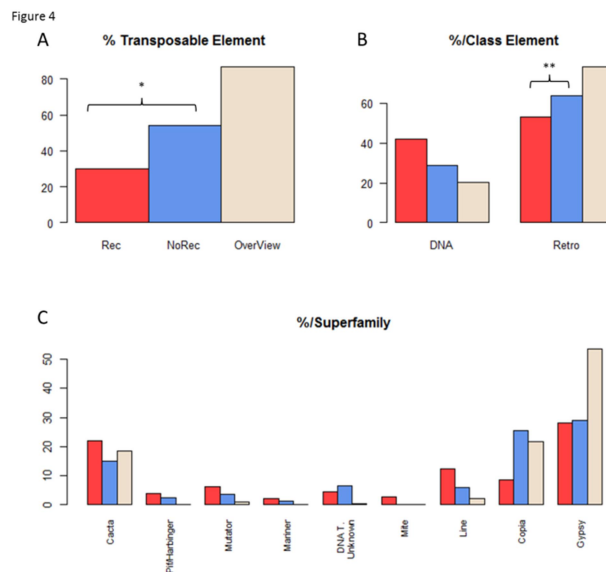
### C. COs occur more frequently in the vicinity of the coding fraction of the chromosomes

For each of the 7,264 predicted genes described in [121], we estimated potential correlations between COs and the presence of genic features (promoter, gene body, terminator) in our 74 Rec intervals (see M&Ms for details). Because of their resolution and their size, some intervals may have a unique genic feature (promoter or gene body or terminator) while others may have two or three (see Fig S3 for details).

Among the 252 COs mapped, 179 (74%) were comprised in intervals presenting at least one genic feature. To correlate recombination rate with sequence features, a control set of 74 intervals where no CO was detected (called NoRec hereafter) was defined. To avoid biasing the control set, NoRec intervals were selected in the same scaffolds, in the vicinity of Rec ones and with a similar size.



**Fig 3.** Comparative analysis for gene body and expression between Rec, NoRec and OverView intervals. Repartition of Gene body (A) and their expression during meiosis (% of genes expressed; FPKM  $\geq$  1; B) in Rec, NoRec and OverView (3B chromosome) intervals. \*:  $p < 0.05$ ; :.  $p < 0.1$



**Fig 4.** Percentages of transposable elements in Rec, NoRec and OverView intervals for chromosome 3B. Percentages: of transposable elements (A), of TEs according to their classes (B; DNA transposons (DNA) and Retrotransposons (Retro)) and of TEs according to super-families (C). Data are extracted from chromosome 3B pseudomolecule: Recombinant (Rec; Red), Not-Recombinant (NoRec; Blue) and chromosome overview (OverView; white) intervals respectively.

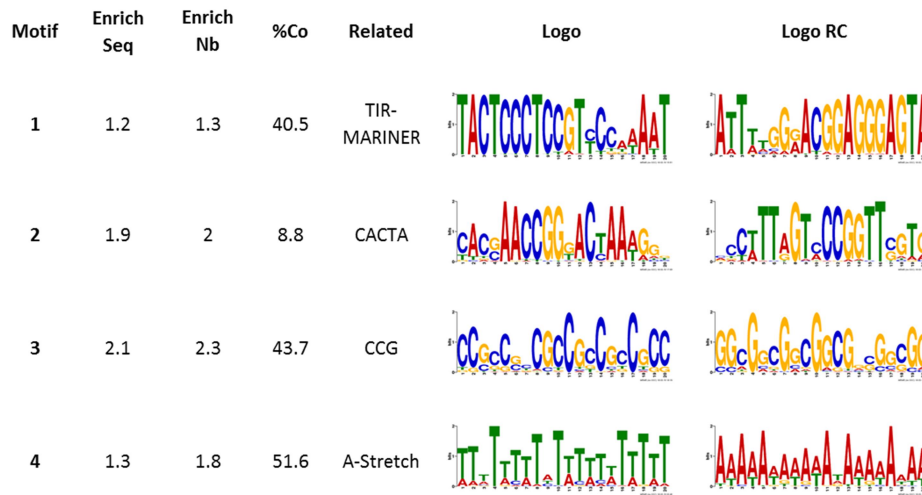
In addition, we created an additional set of 74 control intervals homogeneously distributed along chromosome 3B (hereafter named “chromosome overview”) by selecting each ~10 Mb, 9,344°bp of sequence with no recombination on 406 CsRe lines. The Rec intervals were then compared with the NoRec and chromosome overview (Fig 3). Genic features were present in 73% of the Rec intervals while they were present in only 54% of the NoRec intervals (Fig 3A), a difference which was significant (p-value < 0.05). In addition, expressed genes (FPKM> 1 during meiosis) represented 67% of the genes located in Rec intervals vs 48% for NoRec intervals (p-value < 0.1) (Fig 3B). Chromosome overview intervals showed lower level of genic features (23%) and similar gene expression (50%) than NoRec intervals.

According to our results, recombination in wheat seems to occur more frequently close to or within the genes as in *Arabidopsis* and *C. cerevisiae* [77,296,297]. Recombinogenic regions also carry genes which are more prone of being expressed during meiosis while on the contrary, cold regions carry fewer and less expressed genes. Furthermore, when we achieved a sufficient resolution, 8, 5 and 9 COs were comprised in intervals where only one promoter, gene body or terminator was present respectively, indicating a tendency of recombination to occur more frequently in promoter and terminator. This supports thus the hypothesis that chromatin structure could partly drive CO location [63].

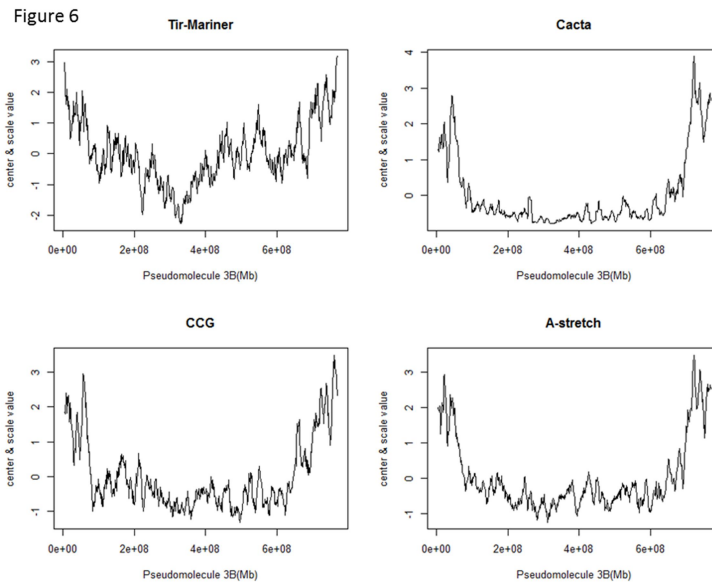
#### **D. Recombinogenic regions have different composition in transposable elements and associated specific DNA motifs.**

Rec intervals contained a significantly lower density of TEs (30%) than NoRec intervals (54%) (p-value <0.05; Fig 4A). Regarding TE composition, Rec intervals carried a similar amount of retrotransposons and DNA-transposons (42% and 53%, respectively) contrary to control intervals which contained 3-4 times more retrotransposons (63% and 78%) for NoRec and chromosome overview intervals, respectively, than DNA-transposons (28% and 20%, respectively), highlighting a particular signature of loci where COs take place.

Figure 5



**Fig 5.** Motif detection on Rec sequence using MEME suite. Only the four more frequent motifs are mentioned. Enrich Seq: ratio between Rec and NoRec intervals (nb intervals/nb intervals); Enrich Nb: ratio between Rec and NoRec intervals (nb motif/nb motif); %CO: % CO associated to the motif; Related: annotation of the motif; Logo: 5'-3' sequence; LogoRC: reverse-complement sequence.



**Fig 6.** Distribution of the four most frequent motifs along 3B pseudomolecule (in bp). Distribution is estimated along chromosome 3B pseudomolecule (774 Mb) using a sliding window of 10 Mb and a step of 1 Mb.

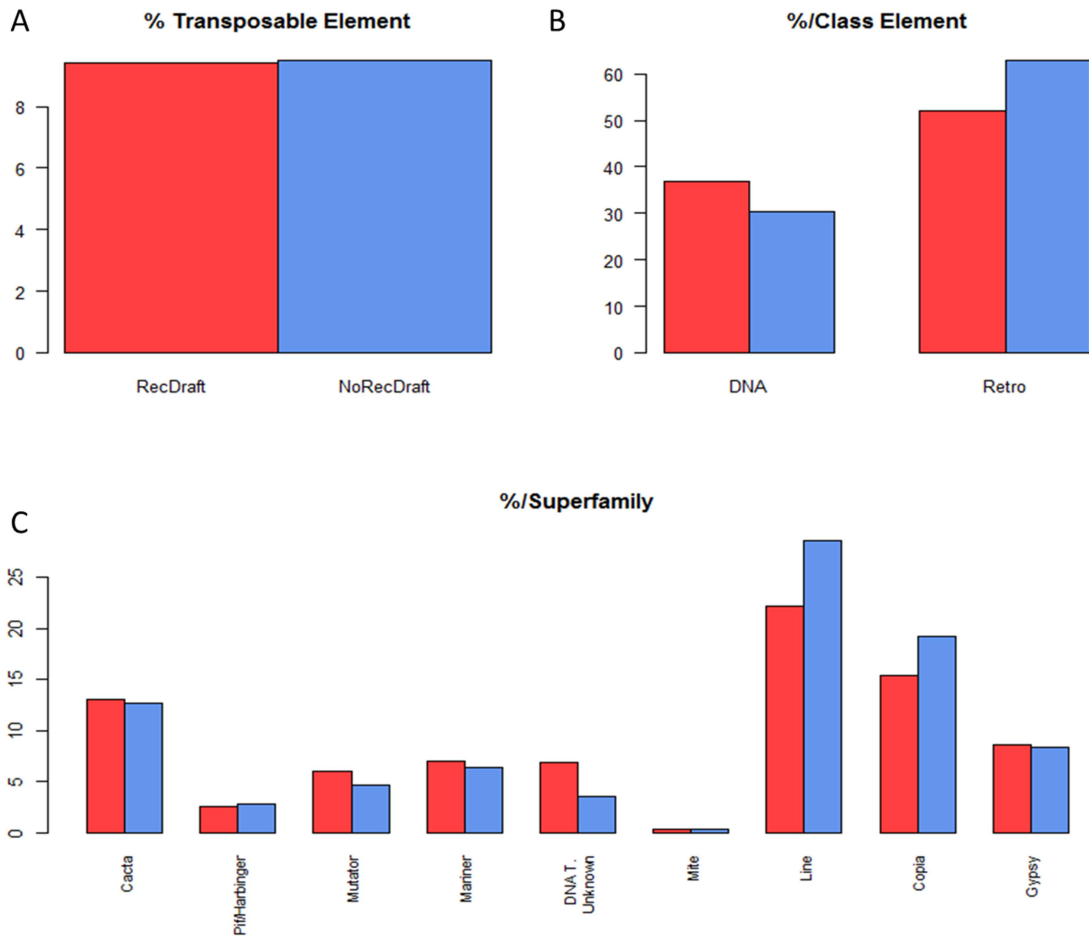
In addition, NoRec intervals showed a significant higher level of Retrotransposons compared to Rec intervals (p-value <0.05, Fig 4B). At the superfamily level (Fig 4C), *CACTA*, *PIF/HARBINGER*, *MUTATOR*, *MARINER*, *MITE* and *LINE* were more frequently observed in Rec intervals, while *GYPHY* and *COPIA* were prevalent in NoRec intervals.

Previous studies in *Arabidopsis*, *Drosophila* and dogs discovered simple recombination-associated motifs like A-stretch motif, CCT, CCN repeats and CpG [95,298,299]. To evaluate the existence of recombination-associated motifs in wheat, we searched for DNA motifs overrepresented in Rec intervals without *a priori* using MEME suite and base frequency of chromosome (see M&Ms for details). Subsequently, enrichment in Rec vs NoRec intervals was evaluated as well as percentage of COs present in sequences which carry these motifs. Distribution along 3B pseudomolecule was also analyzed. Sixteen motifs were identified in sequences surrounding COs (Fig S4). We focused our analysis on four of them because they were found between 1.2 to 2.1 times more frequently and showed a higher copy number in Rec compared to NoRec intervals (Fig 5).

Two motifs are simple repeats: [CCG]*n* (p-value = 0.187) and poly-A stretches (p-value = 0.145); two are parts of known TEs: CTCCCTCC (p-value = 0.527) in terminal inverted repeats of *TC1/Mariner* superfamily [300] and TTAGTCCCGGTT (p-value = 0.362) found in *CACTA* elements. All four motifs followed recombination pattern and partitioning of chromosome 3B with higher proportion in the two distal regions (Fig 6; all motifs in Fig S5). Each motif was present on Rec intervals which carry 8.7 to 51.6% of the COs in our dataset (Fig 5).



Figure 7



**Fig 7.** Percentages of transposable elements in RecDraft and NoRecDraft intervals for the whole genome. Percentages: of transposable elements (A), of TEs according to their classes (B; DNA transposons (DNA) and Retrotransposons (Retro)) and of TEs according to superfamilies (C). Data are extracted from the whole genome: Recombinant (RecDraft; Red) and Not-Recombinant (NoRecDraft; Blue) intervals respectively.

### E. Data from chromosome 3B can be expended to the whole genome

Whole-genome genetic map [291] allowed us to analyze recombination on 406 F6 lines of CsRe which led to validate 596 COs in 476 intervals of less than 26 kb (called RecDraft) (Table S2). Ninety percent of COs were delineated in intervals of less than 10 kb with 81% of intervals presenting only 1 CO. As for chromosome 3B, we selected control intervals devoid of recombination (NoRecDraft). Similar results as those observed for chromosome 3B were found: more genes in RecDraft intervals (66% vs 58% in NoRecDraft) and a tendency of these genes to be expressed during meiosis (73% vs 68%) (Table 1). The percentage of TEs was close to 10% in both intervals of the draft genome (Fig 7A). Nevertheless, same tendency as for chromosome 3B was observed with more DNA transposons in RecDraft intervals and more Retrotransposons in NoRecDraft intervals (Fig 7B). However, analysis of TE composition revealed some similarity with chromosome 3B like more *Mutator* elements (+1,4%) in RecDraft intervals and some difference like more *LINE* and *COPIA* elements in NoRecDraft intervals (Fig 7C).

Comparative analysis between results obtained on 3B pseudomolecule and draft genome sequence of bread wheat. Gene information is provided by presence of CDS in intervals and expression by FPKM > 1.

Associated recombination motifs were search on RecDraft intervals with the same approach as the one applied on chromosome 3B. Ten motifs were found among which, seven were related to the first 20 base pairs of the *TIR-Mariner* superfamily (Fig S6). [CCG]<sub>n</sub> was present within intervals of RecDraft where 16% of COs were mapped (Table 2). The motif related to *TIR-Mariner* as well as A-stretch were present within intervals which carried 26% of COs in RecDraft. Enrichment and copy numbers were close to those observed for chromosome 3B for the motif related to *TIR-Mariner* and slightly lower for [CCG]<sub>n</sub> (Table 2) while they were lower for the motif related to *CACTA* and A-stretch.



Comparative analysis of recombination-associated motifs detected on 3B pseudomolecule and on Draft Genome. Enrich Seq: ratio between Rec and NoRec intervals (nb intervals/nb intervals); Enrich Nb: ratio between Rec and NoRec intervals (nb motif/nb motif); %CO: mean % CO associated to the motif.

In the light of the results and despite the bias generated by the size and content of the contigs, we confirmed our previous observations from chromosome 3B with the same proportions of genes and transposable elements in RecDraft and NoRecDraft. We had the same trend regarding gene and gene-expression in meiosis and RecDraft intervals are enriched in DNA transposons while NoRecDraft intervals are enriched in retrotransposon. Concerning the motif discovery, the related-to *TIR-Mariner*, A stretch and CCG motifs were confirmed putatively related to recombination in the whole-genome analysis albeit the lower enrichment factor and copy number.

## 7. Discussion

### A. New wheat-genomic tools allow more resolute analysis of recombination pattern

Whole-genome fine-scale recombination studies in plants have only been conducted in species for which the genome sequence was available [67,88,95,102,284]. However, this restricted such analyses to species with small and compact genomes with low rates of transposable elements. Only maize had a large genome and data showed a correlation between *Mu* transposable element insertion sites and CO localization [88]. Bread wheat has a six-time larger genome than maize and recombination at the chromosome-scale is low and occurs preferentially in distal regions. It is thus positively correlated with gene density and negatively with transposable element content [121].



We exploited the pseudomolecule of chromosome 3B [121] as well as high throughput detection of SNPs [291] to map 252 CO events in intervals of less than 26 kb using genotyping data. This resolution has never been reached in wheat before and was close to the one observed for the four hotspots described in rice [67]. In the collared flycatcher, a songbird species, 279 COs were localized within intervals <10kb [82] which was only 1.8 times more compared to our analysis (156 COs) while the wheat genome is 17 times larger. Similarly in Human, genome wide sperm-sequencing allowed mapping of 13-45% of CO in 30 kb intervals [301,302] showing that even in a highly documented genome 6-fold smaller than that of wheat, reaching a resolution lower than 30 kb is difficult. Such a deep analysis has never been conducted in wheat before due to the large genome size (~17Gb) and the high proportion (~85%) of repeated elements. 252 COs analyzed seems a few compared to the 3,031 detected in the whole population on chromosome 3B. However, the remaining 2,779 were mainly present between or on large scaffolds preventing their resolution and despite development of numerous SNPs.

#### **B. Ancestral and current recombination follow the same pattern**

In avian [82,139] and in *Saccharomyces* species [140], recombination patterns are highly conserved during evolution. This conducted us to assay another approach, to reveal historical recombination using linkage disequilibrium and coalescence theory (Rho-map [281,292]) in wheat. Our results revealed that the hotspots observed in collections are present in the F6 population, confirming their stability across many generations. In our study, the choice of the genotypes could be discussed because it presented the maximum divergence between the genetic pools and selection pressure can be present, especially in the European genetic pool. Then, lack of resolution in some regions could influence the estimation of recombination intensity ( $\lambda$ ) and the presence/absence of SNP markers can suggest some structural polymorphisms which could also influence the estimation of historical recombination rates.



However, common break points were found between the two genetic pools and CsRe population. This demonstrated that this strategy could be used with the data already existing to define complete recombination-hotspot maps of bread wheat as this was done in Human [85]. Specific recombination intensities were different depending on the collection used which could be attributed to the difference in LD patterns at fine scale in the two populations, difference arising from the complex evolutionary histories of the two genetic pools we used.

This first approach offers opportunity to use recent high-throughput genotyping data [291] but in the future, massive data such as those generated by genotyping by sequencing (GBS; [303]) can lead of complete recombination map in wheat, allowing the same approach of parameter detection as realized in Human and *Arabidopsis* [77,86,87]. We could thus identify recombination breakpoints located in pericentromeric regions which could be enhanced to improve recombination in these poorly-recombinogenic regions.

### C. Retrotransposons reduce recombination rate

We found that NoRec intervals have significantly higher TE content (53%) and more retrotransposons (*Gypsy* and *COPIA* elements; 63.7%) than Rec intervals. Interestingly, the percentage of *COPIA* element is even higher in NoRec intervals than in the central part of chromosome 3B (called R2 in [121]) which is made of 70% of retrotransposons and exhibits very low recombination rate (0.05 cM/Mb; [121,123]).

The negative impact of LTR transposons on recombination in wheat could be attributed to their associated-epigenetic patterns which condenses DNA and locks the regions preventing their accessibility to the recombination machinery. As an example in mouse, methylation leads to DNA conformation which prevents transposons from adopting permissive chromatin structure for meiotic recombination [93].





Finally in barley, pericentromeric regions with low recombination rates bear LTR-retrotransposons that carry the epigenetic landmarks H3K9me2 and H3K27me1 which establish a constitutive heterochromatic state [97]. These results support the hypothesis that crossovers do not occur – or only very rarely – in chromosomal regions where chromatin is condensed and poorly accessible to recombination mechanism.

#### **D. Recombination and DNA-transposons are located in the same regions**

Analysis of data at the chromosome 3B and whole-genome scales showed that recombination more often occurs in or close to genes that tend to be expressed at meiosis. In total, 75% of the COs from chromosome 3B fall into an interval containing at least one gene feature (promoter, gene body or terminator) confirming thus the results observed in yeast [304]. At the whole-genome level, similar results were found with 2/3 of COs overlapping genes. Our COs were found in the gene body and slightly more frequently in promoters and terminators confirming the results observed in *Arabidopsis* [77] as well as in avian [82,139] and in *Saccharomyces* species [140], where recombination pattern is highly conserved and localized to the promoter regions.

We also found higher rate of DNA transposons – albeit not significant– in Rec intervals. This could be explained by a higher survival rate of DNA transposons inserted close to genes and a remaining opportunist and strategic processes to maintain – or even under certain conditions to increase – the number of DNA transposons. An example in maize was described at the *al* locus where *MuDR*, autonomous element of *Mutator* family which encodes a transposase, supported the capacity of DNA transposons to influence recombination [305].



Epigenetic studies showed that H3K4me3 and H2A.Z landmarks are important for promotion of gene transcription [77,78,88] and are associated with high recombination rates in Arabidopsis [78] as well as in barley [97,98]. This suggests that both recombination and DNA-transposon insertion benefit from low compaction of DNA to occur within the genome.

#### **E. TE-related motifs are thus present in recombinogenic windows**

We revealed four motifs on chromosome 3B and at the whole genome level that were overrepresented in Rec intervals: two simple motifs (A-stretch and [CCG]) and two related to transposable elements (*TIR-Mariner* and *CACTA*). Research along the pseudomolecule of chromosome 3B for the various motifs showed a similar partitioning as recombination and preferential location of these four motifs in distal regions of the chromosome.

These two different categories (complex TE-associated motif and simple repeats) were found in previous analyses in other species [77,95,298,299,306]. A-stretch motif was already found in Arabidopsis, Human and Drosophila as associated with recombination in ~2 kb areas around CO events suggesting that A-stretches do not directly induce recombination but that they contribute to its occurrence as a result of DNA de-compaction [86,95,298]. A-stretches were described as likely stiff in order not to fold around nucleosome and are thus a strong inhibitor to nucleosome formation. Presence of A-rich motifs close to genes could regulate their expression [96] and may thus reflect a higher level of expression and a less condensed DNA.

The second motif found in wheat is CCG that could be related to CCN found in canids, Arabidopsis and Drosophila [95,298,299]. Our complete motif was CCGCCGNCGCCGCCGCCGCC. It shares similarity with CCNCCNTNNCCNC motif found in Human hotspots and corresponding to the *PRDM9* site [86,307]. CCN motif is associated with H3K4me3 epigenetic landmark and the gene body of the genes in Arabidopsis and could thus influence chromatin compaction [95].



For the first time, related *TIR-Mariner* motifs are putatively associated with recombination. *TC1-Mariner* element is the most diverse group and widespread super-family in eukaryotes and it was extensively studied in animals [308]. Its insertion site is in TA sequence and like DNA transposon, it generally locates in genic regions [300]. In our case, most of the related *TIR-Mariner* motifs detected that are present in Rec intervals are found close to gene features (73% for chromosome 3B).

#### F. *TIR-Mariner* motif is highly similar as those described for *PRDM9*

The related *TIR-Mariner* superfamily motif tends to be more present in Rec intervals that we defined on chromosome 3B and at the whole genome scale. Interestingly, part of our related *TIR-Mariner* motif (TCCCTCC) is similar to the *PRDM9* motif (CCTCCCT) driving DSB location in Human [309] and contains five common bases (TCCCT). CCTCCCT motif is more frequent in Human hotspots and is overrepresented (5-6 fold) in primate specific retro transposon *THE1A/B* consensus [86,87]. *PRDM9* leads recombination by ensuring H3K4me3 methylation [48]. It seems to be very specific to a few species and is rather rare among eukaryotes. *PRDM9* is absent in plants [94], birds [310], *Drosophila* [311] and even in some mammals such as the canids [299,310]. On the contrary, recombination is essentially universal.

In mouse *prdm9*-mutants, recombination hotspots are reverted to gene promoters as it is observed in yeast, plants and in our analysis suggesting a functional gain to specify location of recombination [46,78]. In canids where *PRDM9* is absent, recombination occurs in CpG-rich regions around promoters with little association with H3K4me3 marks [299]. These observations suggest a common ancestral regulatory system to control recombination involving transposable element such as the *THE1A/B* in Human or the *Mariner* elements in plants. TIRs are essential for transposition since they are the transposase binding sites. Thus transposase-mediated DSBs may initiate recombination as *SPO11*-directed DSBs at meiosis.



Either, these TEs could be in specific conformation which allow the fixation of host machinery to create DSBs as this was shown in Mouse [93]. We can hypothesize that in a few mammals such as Human, the system has evolved and *PRDM9* was set up to control breakage and recombination. In plants, there are many difficulties to find a predominant mechanism to explain recombination localization and probably multiple mechanisms coexist. This could lead to more plasticity increasing fitness in plants to ensure adaptation in changing environments.

## G. Conclusion

Until now, no study regarding recombination hotspots was led at this scale and resolution in bread wheat. With a large segregating population (1270 individuals), we were able to map about 900 CO events using latest genomic sequences available. Availability of the full genome sequence and use of another population will allow determining more precisely hot and cold regions in bread wheat because in the light of previous studies, we obviously did not capture all the recombination breakpoints and ancestral recombination could be an excellent alternative since we showed that the correlation with current recombination is elevated especially when the resolution is high.

Recombination occurs more frequently in regions associated with genes, especially with promoters or terminators. These regions are also enriched in DNA-transposons and especially from the *Mariner* family. We identified a correlation between recombination events and a motif associated with the *TIR* sequence of *Mariner* similar to the *PRDM9* motif derived from *THE1A/B* element. We can thus hypothesize that ancestrally, recombination and DNA-transposon (mostly *Mariner*) insertion occurred in regions owning a structural opened chromatin. This strategy has been conserved in most eukaryotes except a few like Human where the *TIR* part of the element has been alienated to direct double strand breaks. Our study performed without *a priori* in bread wheat is in favor of a common general regulatory system of recombination distribution.



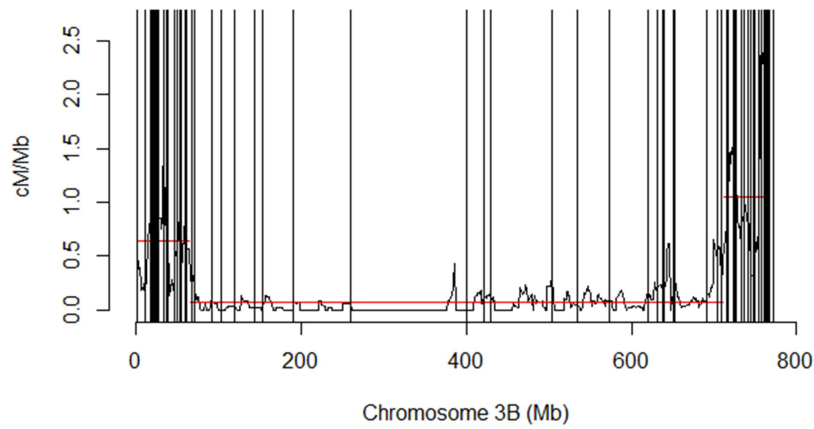


## Acknowledgments

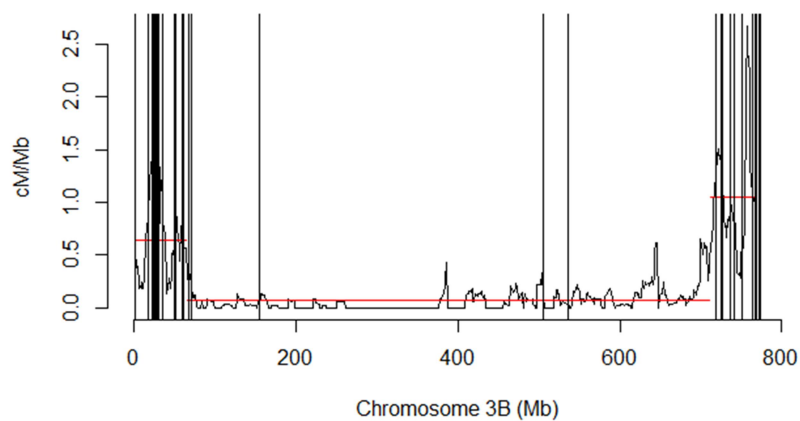
The authors are grateful to Alain Loussert who performed all the genotyping experiments as well as to the personal from GENTYANE for technical help. The authors also thank Christine Mézard and Ian Henderson for their help during redaction and for discussions. The seeds of both Asian and European accessions were provided by the Biological Resources Center for small grain cereals (INRA Clermont-Ferrand). BREEDWHEAT project and IWGSC consortium are also greatly acknowledged for providing SNP (Axiom) genotyping data and wheat genome sequences respectively. BD was supported by INRA (meta-program SELGEN) and Region Auvergne.



## Supporting Information

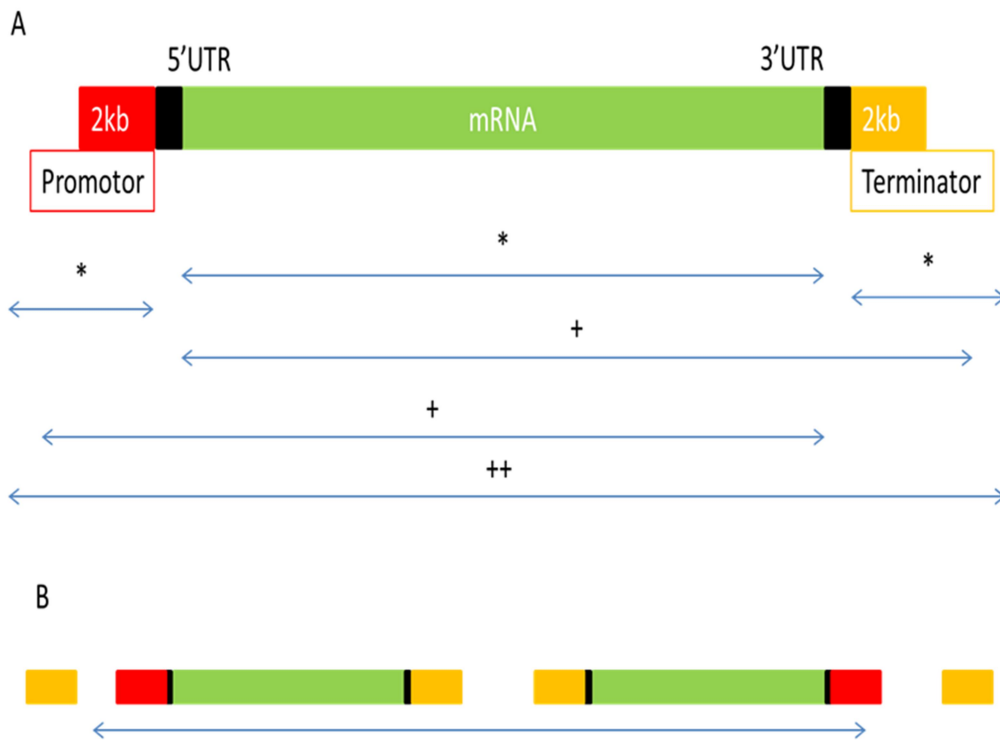


**Fig S1.** Localization of the 73 scaffolds with recombination detected along the 3B pseudomoleculeV6. Distribution of the recombination along the 3B pseudomoleculeV6 (in red, sliding window 10Mb, step 1Mb). Each bar represents the starting point of the scaffold where recombination was detected.



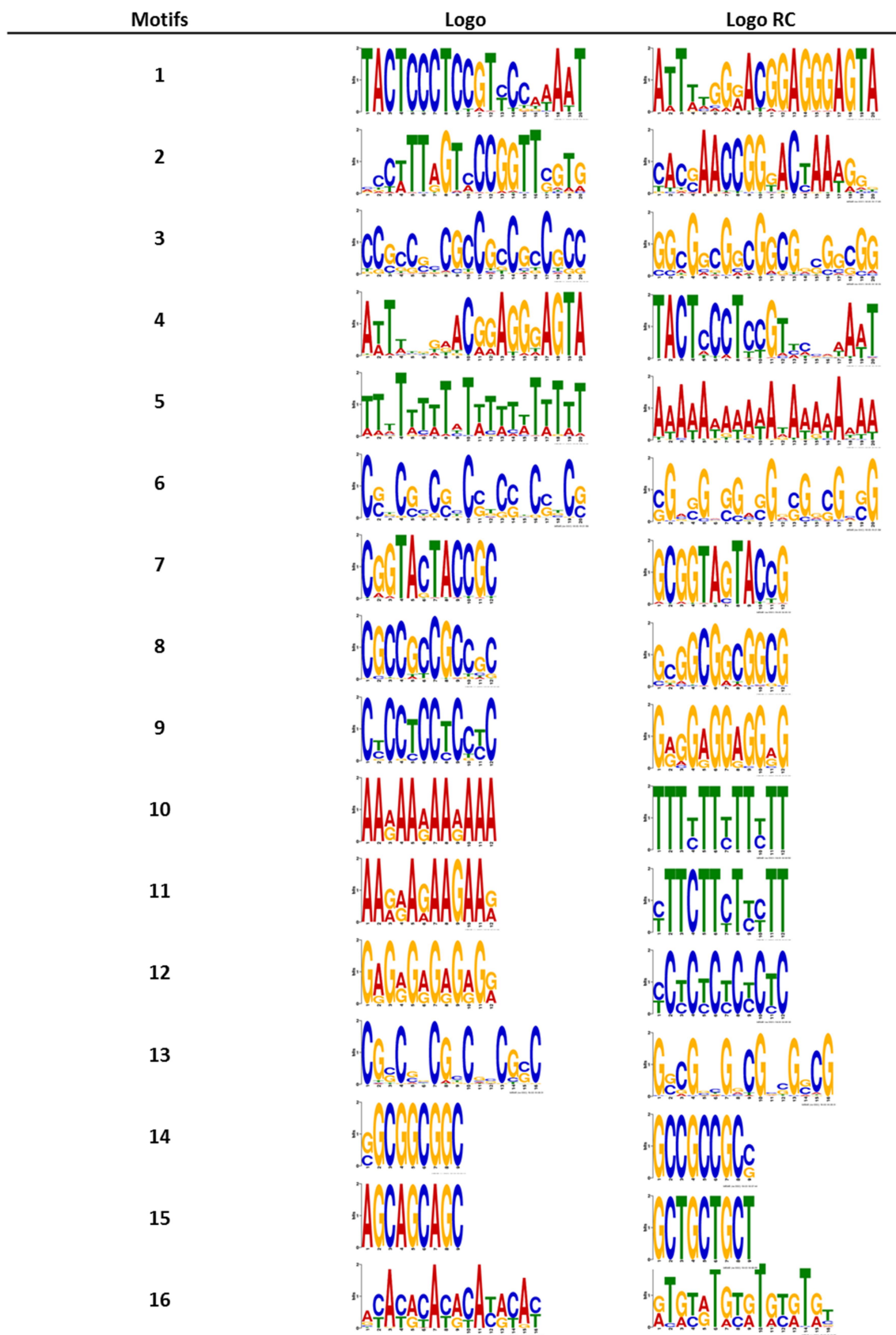
**Fig S2.** Localization of the 74 intervals with recombination detected along the 3B pseudomoleculeV6. Distribution of the recombination along the 3B pseudomoleculeV6 (in red, sliding window 10Mb, step 1Mb). Each bar represents the starting point of the interval where recombination was detected.





**Fig S3.** Location of COs associated with gene models. (A) Gene model is constituted of the promoter (red), the gene body (green) and the terminator (orange). Promoter and terminator are defined as 2kb up or down stream after UTR (black) features. Respective resolution of our intervals is indicated with a blue double arrow: \*: location of COs in only one part of the gene model; +: location of COs in two parts of the gene model (promoter + gene body or gene body + terminator); ++: location of COs in the full gene model (promoter + gene body + terminator). (B) When the resolution was not sufficient, COs may be associated to more than one gene model.

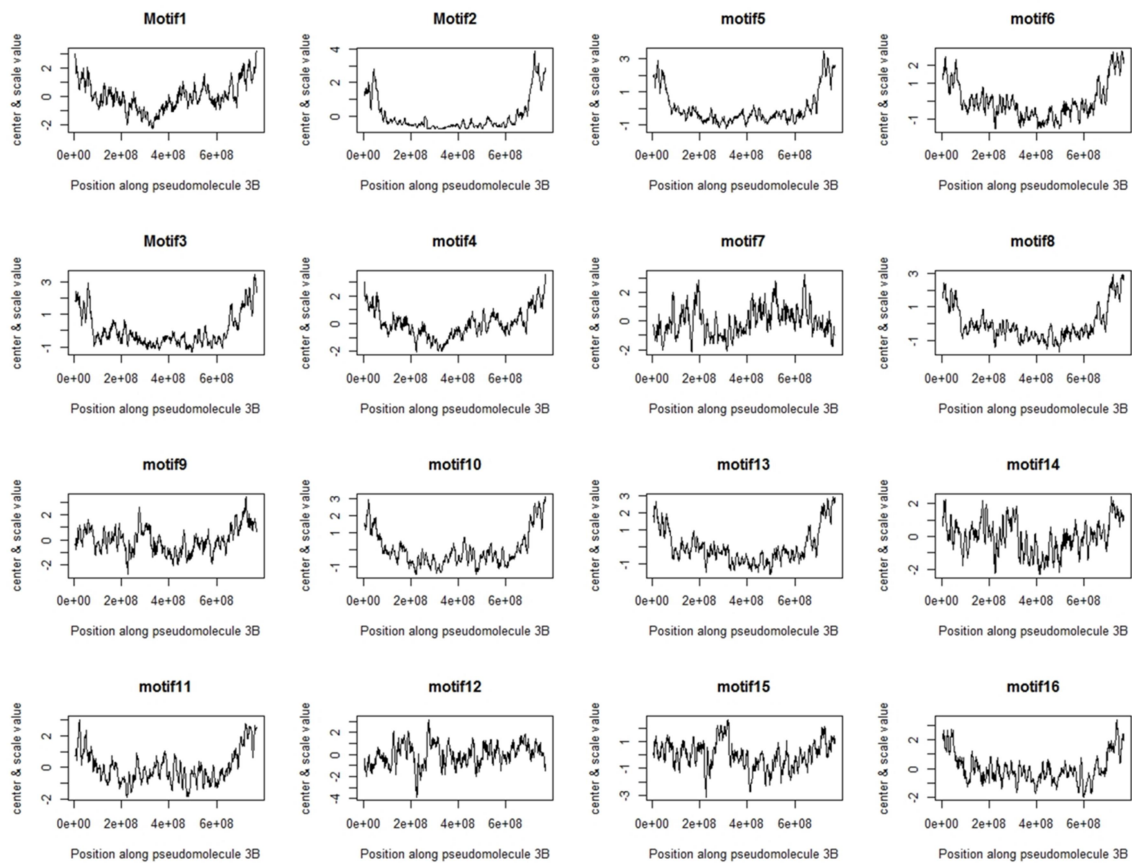




**Fig S4.** List of the 16 motifs detected on Rec intervals.

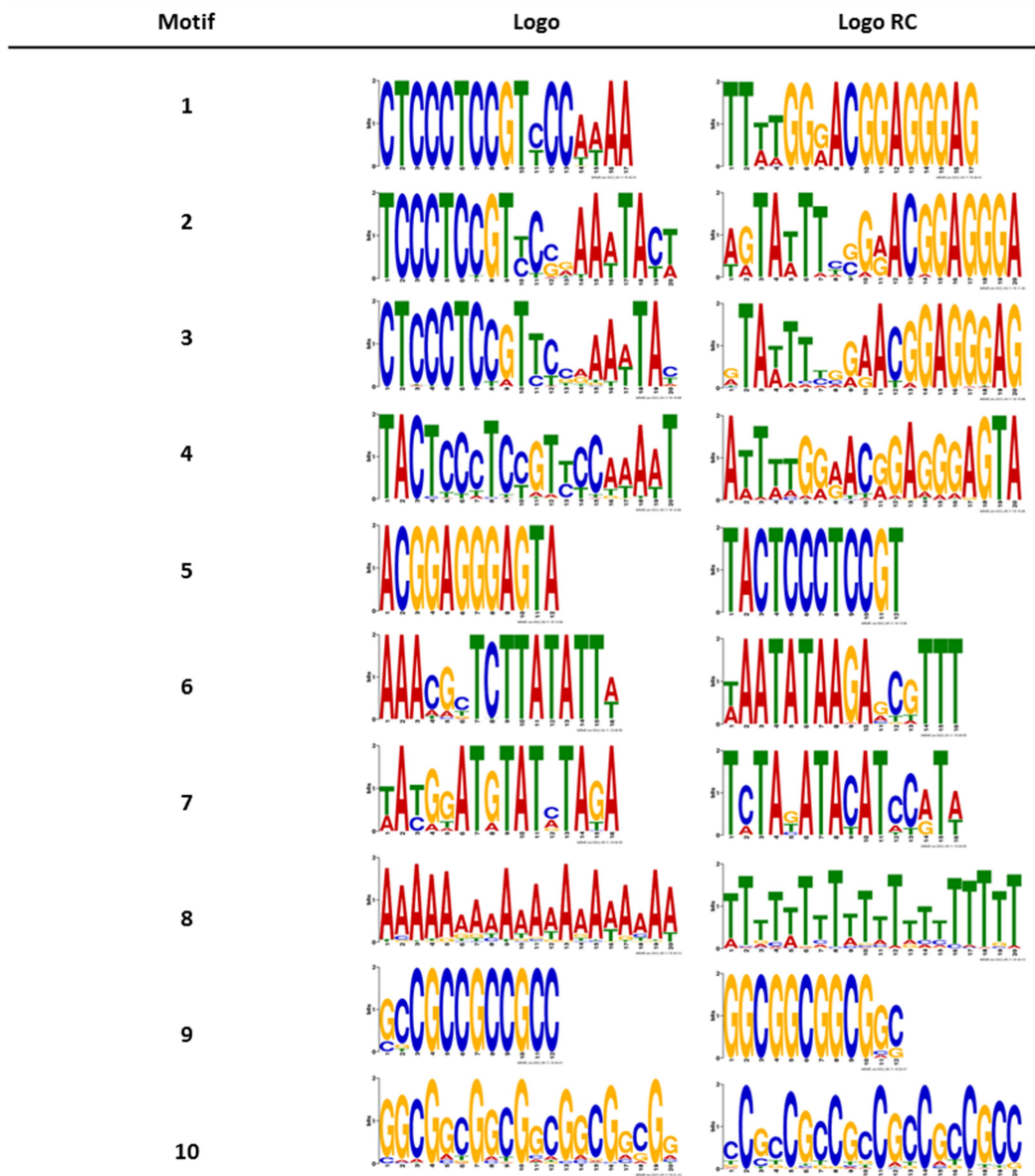






**Fig S5.** Distribution of the 16 motifs along chromosome 3B pseudomoleculeV6. Center & scale values of the number of motifs detected along the chromosome 3B pseudomoleculeV6. Sliding window of 10Mb step: 1 Mb.





**Fig S6.** List of the 10 motifs detected on RecDraft intervals.



**Table S1.** List of the 73 scaffolds with recombination detected along the 3B pseudomoleculeV6. Start and Stop indicate the position of the scaffold along the pseudomolecule.

**Table S2.** List of the 476 contigs from IWGSC survey sequence ([12]) used to detect recombination at the whole-genome scale.

**Table S3.** List of the 180 accessions from the two genetic pools (Asian and European) used for ancestral recombination analysis. ERGE indicates the classification number from the Resource Centre.

**Table S4.** List of the 774 SNPs and of the 61 SSRs developed on 3B chromosome for the improvement of the resolution of the intervals containing COs.



## 2. Principales conclusions

Dans cette étude, nous avons affiné la localisation de près de 900 crossing-overs distribués sur la totalité du génome avec une résolution jamais atteinte jusqu'à présent chez le blé. Cela nous a permis de voir que la recombinaison ancestrale déterminée par la variation du déséquilibre de liaison était majoritairement conservée par rapport à la recombinaison actuelle. Cela confirme une évolution des points chauds de recombinaison et suggère que les points ancestraux localisés en région péricentromérique et qui ne sont plus actifs sur notre population pourraient l'être sur une autre ou être réactivés sous certaines conditions (stress, ciblage...).

Nous avons également confirmé que les crossing-overs apparaissent plus fréquemment dans les parties exprimées du génome, et plus particulièrement dans les parties promotrices des gènes. Cela suggère que l'accessibilité de la chromatine joue un rôle prépondérant dans la réalisation des événements de recombinaison. Il serait donc pertinent d'étudier plus en détail les marques épigénétiques qui gouvernent la compaction de l'ADN et d'analyser le comportement méiotique de mutants des gènes qui contrôlent le dépôt de ces marques.

En plus des motifs traditionnellement associés aux crossing-overs, nous avons trouvé des motifs apparentés aux parties terminales (TIR) des éléments transposables de type Mariner. Ces motifs partagent des similarités avec le motif découvert chez l'Humain et lui aussi associé à un élément transposable et contre lequel est dirigée la protéine PRDM9. Cela suggère que les éléments transposables et la recombinaison apparaissent dans les mêmes sites à l'issue de cassures double-brin et que chez un nombre limité d'espèces, cet aspect a pu être détourné sous la forme de la protéine PRDM9 qui est apparue soit pour cibler plus facilement les régions de cassures et améliorer la recombinaison, soit pour se protéger des éléments transposables en les scindant.





## V. Conclusion générale et perspectives

---

L'amélioration des rendements dans un contexte d'agriculture intensivement durable est un enjeu majeur actuel pour permettre la sécurité alimentaire. Afin d'y parvenir il est possible et nécessaire de continuer l'effort de sélection et d'utiliser au maximum la réserve d'allèles originaux que sont les ressources génétiques. Cependant, afin d'utiliser de la meilleure manière et le plus rapidement possible cette ressource, il est nécessaire de mieux comprendre où et comment se passe la recombinaison méiotique chez le blé tendre et chez d'autres espèces cultivées. Dans le cadre de ma thèse, j'ai développé des approches de cytogénétique, de biologie moléculaire, de génétique et de bio-informatique afin de contribuer à ces connaissances chez le blé tendre.

Ainsi, il a été possible de transférer les connaissances cytogénétiques disponibles chez l'orge vers le blé. Grâce à ces techniques particulières d'observation en 3D des méiocytes j'ai pu observer le déroulement de la méiose chez le blé tendre et la comparer avec l'orge. Ces comparaisons ont permis d'approfondir les travaux d'analyse des gènes du chromosome 3B qui impactent le comportement méiotique.

Grâce à ces images obtenues chez les lignées délétées de tout ou partie du bras long du chromosome 3B et qui présentaient un nombre de chiasmas inférieur à la plante sauvage, j'ai pu montrer que le synapsis avait bien lieu chez ces mutants contrairement à ce qui avait été décrit précédemment dans la littérature. Nous avons ainsi pu observer qu'au moins deux gènes localisés sur ce bras de chromosome influent sur le nombre de chiasmas et grâce aux analyses bioinformatiques et les données récentes de génomique, j'ai proposé quelques gènes candidats qu'il sera nécessaire de valider.



L'étude des gènes méiotiques a très largement porté sur *Arabidopsis* et les techniques de cytogénétique ont été très largement utilisées pour les caractériser et observer leur action lors de la méiose. Chez le blé ces techniques se développent et ont déjà permis de répondre à certaines questions concernant le contrôle de l'appariement homéologue par *Ph1* [130]. Le développement des techniques d'immuno-histologie en trois dimensions sont prometteuses et l'avancée des technologies de microscopie permet aujourd'hui une visualisation du très petit à des résolutions inégalées. L'utilisation de telles techniques chez le blé tendre devraient permettre de répondre aux questions spécifiques associées à cette espèce et plus largement, permettre de trouver les inter-acteurs de la recombinaison en utilisant comme base les données déjà acquises sur les espèces modèles diploïdes comme *Arabidopsis* ou l'orge.

Une ressource essentielle pour l'identification des inter-acteurs de la recombinaison est la mise à disposition de banques de mutants. Chez *Arabidopsis*, ce sont les mutants de fertilité qui ont permis l'identification d'un grand nombre de gènes/protéines impliqués dans la recombinaison chez les plantes. Chez le blé de telles ressources sont en cours de caractérisation au sein de notre institut ou du John Innes Centre au Royaume-Uni par exemple. Le développement de mutants de gènes méiotiques chez le blé est essentiel car ceux-ci permettraient à termes de « contrôler » (activer/désactiver) la recombinaison dans les schémas de sélection et ainsi créer plus rapidement des blés adaptés aux enjeux mondiaux.

Cependant, l'état polyploïde du blé tendre, la complexité de la régulation de l'expression des différentes copies homéologues et l'existence de copies issues de duplications restent un défi. Par conséquent le ciblage des copies fonctionnelles et le cumul des mutations pour un même gène ou différents gènes peuvent s'avérer difficiles. La création de telles ressources va exiger plusieurs années.



Néanmoins, les nouvelles technologies d'édition des génomes telles que l'utilisation des systèmes CRISPR-CAS9 [312,313] représentent une alternative qui doit être envisagée même si l'acceptation des techniques d'édition du génome demeure un problème au niveau de la production de nouvelles variétés.

J'ai également réalisé une cartographie haute densité du génome de blé tendre. Disposer d'un très grand nombre de marqueurs moléculaires ordonnés est un atout pour le sélectionneur afin d'appliquer des schémas de sélection génomique et optimiser ainsi les parcelles d'essais. C'est également d'un grand intérêt pour l'assemblage et/ou la vérification du génome, les études de génomique structurale et en particulier pour avoir une vision fine des profils de recombinaison au niveau de l'ensemble du génome. J'ai ainsi pu montrer que le partitionnement observé sur le chromosome 3B est généralisable à tous les chromosomes du blé tendre et que ce chromosome particulier constitue donc un excellent modèle d'étude.

L'utilisation de modèles en biologie est courante afin de réduire la complexité avant d'intégrer ces données comme base de compréhension sur les espèces d'intérêt et de développer les outils nécessaires. Ainsi la compréhension des génomes complexes des céréales a débuté par des études exhaustives et le séquençage des génomes plus simple d'*Arabidopsis*, du riz ou de *Brachypodium*. Si les informations produites ont été capitales pour une meilleure connaissance des fonctions des gènes – et en particulier des gènes méiotiques dont le fonctionnement semble universel chez les eucaryotes – les données d'analyses structurales sont beaucoup moins convaincantes du fait de la différence colossale entre ces génomes et celui du blé, polyploïde et envahi d'éléments répétés.

Cependant, ces dernières années bon nombre de verrous technologiques ont été levés et l'étude des génomes complexes est désormais possible grâce aux mêmes outils que pour les espèces modèles.



J'ai également pu voir que les profils de recombinaison sont identiques entre des chromosomes homéologues et donc que les facteurs qui gouvernent la distribution de la recombinaison sont localisés dans les parties conservées entre les chromosomes c'est-à-dire préférentiellement les séquences codantes.

Le cœur de mes travaux de thèse a porté sur l'identification des caractéristiques de la séquence susceptibles d'expliquer la localisation des points chauds de recombinaison ainsi que la répartition du déséquilibre de liaison le long du chromosome 3B, chromosome modèle du blé tendre.

J'ai ainsi montré que les points de recombinaison ancestraux identifiés par une analyse de coalescence et ceux actuels observés dans une population de lignées recombinantes, sont majoritairement localisés aux mêmes endroits. Cela suggère que chez le blé, les points de recombinaison évoluent peu contrairement aux mammifères comme l'Humain ou la souris [48]. Une analyse de coalescence génome entier est donc envisageable afin d'identifier des points de recombinaison dans les régions péri-centromériques faiblement recombinogènes. Ces points pourraient être actifs dans d'autres populations ou pourraient peut-être être réactivés par des stratégies de ciblage de la recombinaison telles que celles développées chez la levure [314].

Mon étude montre que la recombinaison chez le blé tendre se situe dans le voisinage des séquences codantes et plus particulièrement dans les parties promotrices des gènes. Cela suggère que la recombinaison profite d'un environnement chromatinien ouvert pour se réaliser et que le blocage observé dans les régions péri-centromériques est peut-être dû à une compaction trop grande de l'ADN. Des études conduites sur des mutants structuraux de l'ADN (méthylation) montrent que le profil de recombinaison peut être modifié [288,315].





Par ailleurs, une variation de la température chez l'orge lors de la méiose modifie également le profil en déplaçant la recombinaison vers le centromère [316]. Cependant, ces mutants ou ces traitements induisent simultanément une stérilité qui est un handicap au moment de la sélection. Du fait de son statut polyploïde, le blé serait peut-être plus tolérant à ces mutations et l'étude de mutants de marques épigénétique reste une piste intéressante pour améliorer la recombinaison dans les régions péri-centromériques habituellement peu recombino-gènes. Chez le blé tendre nous ne disposons pour l'instant que de très peu d'informations sur les marques épigénétiques et à une résolution ne permettant que très peu d'interprétation. Ainsi, seules la séquence et l'expression nous permettent de postuler sur l'état chromatinien. Ces données mériteraient d'être acquises rapidement.

Ces régions chromatinienne s'ouvertes semblent attirer de manière préférentielle certains types d'éléments transposables notamment les éléments de type Mariner ou Mutator (et de manière plus générale les DNA-transposons autres que CACTA). Aussi, ces DNA transposons sont présents selon un gradient entre les génomes (D>B>A) [143]. De plus mon analyse des séquences a mis en lumière un motif particulier relié aux éléments de type Mariner et partageant des similarités avec le motif ciblé par la protéine PRDM9. Le motif reconnu par PRDM9 est associé au retro-transposon THE1-A/B, suggérant ainsi que la machinerie de recombinaison a pu être utilisée comme système d'inhibition des éléments transposables au travers de PRDM9 et que cette stratégie pourrait être latente chez les plantes. Identifier un motif particulier plus fréquent dans les régions recombino-gènes est intéressant dans la mesure où ce motif se retrouve ailleurs dans le génome et notamment dans des régions qui ne recombinent pas (ou peu). Il est donc envisageable d'utiliser cette cible toujours avec l'objectif de pouvoir diriger la recombinaison aux endroits voulus.



Augmenter la recombinaison et modifier sa distribution chez le blé tendre reste un enjeu important afin de permettre la séparation de liaisons entre gènes d'intérêt et caractères délétères et ainsi éviter l'autostop génétique. Il est impératif d'identifier les facteurs qui contrôlent ces deux paramètres afin de pouvoir exploiter au mieux les ressources génétiques pour répondre aux besoins d'amélioration des rendements du blé dans un contexte d'agriculture durable.

L'obtention d'une nouvelle variété élite par l'intégration d'allèles nouveaux et performants issues de variétés exotiques qui ne présentent aucun caractère agronomique nécessite un contrôle optimale de la recombinaison entre les génomes des variétés élites et ceux des espèces ou variétés exotiques. Ce contrôle est nécessaire pour répondre aux enjeux mondiaux et à l'augmentation et l'accélération des besoins. Les résultats obtenus dans le cadre de ma thèse sont un premier pas vers une compréhension exhaustive et un meilleur contrôle de la recombinaison chez le blé tendre.



## VI. Bibliographie

---

1. Lenser T, Theißen G. Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* 2013;18: 704–714. doi:10.1016/j.tplants.2013.08.007
2. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature.* 2002;418: 700–707. doi:10.1038/nature01019
3. Ross-Ibarra J, Morrell PL, Gaut BS. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci.* 2007;104: 8641–8648. doi:10.1073/pnas.0700643104
4. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet.* Nature Publishing Group; 2013;14: 840–52. doi:10.1038/nrg3605
5. Matsuoka Y. Evolution of polyploid triticum wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant and Cell Physiology.* 2011. pp. 750–764. doi:10.1093/pcp/pcr018
6. Renny-Byfield S, Wendel JF. Doubling down on genomes: Polyploidy and crop plants. *Am J Bot.* 2014;101: 1711–1725. doi:10.3732/ajb.1400119
7. Wendel JF. The wondrous cycles of polyploidy in plants. *Am J Bot.* 2015;102: 1753–1756. doi:10.3732/ajb.1500320
8. Cui L. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 2006;16: 738–749. doi:10.1101/gr.4825606
9. Doebley JF, Gaut BS, Smith BD. The Molecular Genetics of Crop Domestication. *Cell.* 2006;127: 1309–1321. doi:10.1016/j.cell.2006.12.006

10. Peng JH, Sun D, Nevo E. Domestication evolution, genetics and genomics in wheat. *Mol Breed.* 2011;28: 281–301. doi:10.1007/s11032-011-9608-4
11. Fu Y-B. Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet.* 2015;128: 2131–2142. doi:10.1007/s00122-015-2585-y
12. Abberton M, Batley J, Bentley A, Bryant J, Cai H, Cockram J, et al. Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnol J.* 2016;14: 1095–1098. doi:10.1111/pbi.12467
13. Fita A, Rodríguez-Burruezo A, Boscaiu M, Prohens J, Vicente O. Breeding and Domesticating Crops Adapted to Drought and Salinity: A New Paradigm for Increasing Food Production. *Front Plant Sci.* 2015;6. doi:10.3389/fpls.2015.00978
14. Doré C, Varoquaux F. Histoire et amélioration de cinquante plantes cultivées. *Savoir Faire INRA.* Paris: Institut national de la recherche agronomique; 2006.
15. Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, et al. Long-term Climate Change: Projections, Commitments and Irreversibility Pages 1029 to 1076. *Cambridge Press.* Cambridge University Press; 2014. pp. 1029–1136. doi:10.1017/CBO9781107415324.024
16. Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. The Molecular Biology of Meiosis in Plants. *Annu Rev Plant Biol.* 2015;66: 297–327. doi:10.1146/annurev-arplant-050213-035923
17. Crow JF. Advantages of sexual reproduction. *Dev Genet.* 1994;15: 205–13. doi:10.1002/dvg.1020150303
18. Miko I, LeJeune L. *Essentials of Genetics.* 2014th ed. Cambridge: MA: NPG Education;

2009.

19. SCHUERMAN D, MOLINIER J, FRITSCH O, HOHN B. The dual nature of homologous recombination in plants. *Trends Genet.* 2005;21: 172–181. doi:10.1016/j.tig.2005.01.002
20. Able JA, Crismani W, Boden SA. Understanding meiosis and the implications for crop improvement. *Funct Plant Biol.* 2009;36: 575. doi:10.1071/FP09068
21. Zickler D, Kleckner N. THE LEPTOTENE-ZYGOTENE TRANSITION OF MEIOSIS. *Annu Rev Genet.* 1998;32: 619–697. doi:10.1146/annurev.genet.32.1.619
22. Zickler D, Kleckner N. Meiotic Chromosomes: Integrating Structure and Function. *Annu Rev Genet.* 1999;33: 603–754. doi:10.1146/annurev.genet.33.1.603
23. Pawlowski WP, Cande WZ. Coordinating the events of the meiotic prophase. *Trends Cell Biol.* 2005;15: 674–681. doi:10.1016/j.tcb.2005.10.005
24. Swanson CP. *Cytology and Cytogenetics* [Internet]. Prentice-Hall; 1957. Available: <https://books.google.fr/books?id=x-dqAAAAMAAJ>
25. Zickler D, Kleckner N. Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harb Perspect Biol.* 2015;7: a016626. doi:10.1101/cshperspect.a016626
26. Phillips D, Nibau C, Wnetrzak J, Jenkins G. High Resolution Analysis of Meiotic Chromosome Structure and Behaviour in Barley (*Hordeum vulgare* L.). Lichten M, editor. *PLoS One.* 2012;7: e39539. doi:10.1371/journal.pone.0039539
27. Koszul R, Kim KP, Prentiss M, Kleckner N, Kameoka S. Meiotic Chromosomes Move by Linkage to Dynamic Actin Cables with Transduction of Force through the Nuclear



- Envelope. *Cell*. 2008;133: 1188–1201. doi:10.1016/j.cell.2008.04.050
28. Muniyappa K, Kshirsagar R, Ghodke I. The HORMA domain: an evolutionarily conserved domain discovered in chromatin-associated proteins, has unanticipated diverse functions. *Gene*. 2014;545: 194–197. doi:10.1016/j.gene.2014.05.020
  29. Altmannová V, Kolesár P, Krejčí L. SUMO Wrestles with Recombination. *Biomolecules*. 2012;2: 350–375. doi:10.3390/biom2030350
  30. Armstrong SJ. Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in *Arabidopsis* and *Brassica*. *J Cell Sci*. 2002;115: 3645–3655. doi:10.1242/jcs.00048
  31. Caryl AP, Armstrong SJ, Jones GH, Franklin FCH. A homologue of the yeast HOP1 gene is inactivated in the *Arabidopsis* meiotic mutant *asy1*. *Chromosoma*. 2000;109: 62–71. doi:10.1007/s004120050413
  32. Ross KJ, Fransz P, Armstrong SJ, Vizir I, Mulligan B, Franklin FCH, et al. Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. *Chromosom Res*. 1997;5: 551–559. doi:10.1023/A:1018497804129
  33. Grelon M. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J*. 2001;20: 589–600. doi:10.1093/emboj/20.3.589
  34. Hufton AL, Panopoulou G. Polyploidy and genome restructuring: a variety of outcomes. *Curr Opin Genet Dev*. 2009;19: 600–606. doi:10.1016/j.gde.2009.10.005
  35. Zhang J, Pawlowski WP, Han F. Centromere Pairing in Early Meiotic Prophase Requires Active Centromeres and Precedes Installation of the Synaptonemal Complex in Maize. *Plant Cell*. 2013;25: 3900–3909. doi:10.1105/tpc.113.117846

36. Zickler D. From early homologue recognition to synaptonemal complex formation. *Chromosoma*. 2006;115: 158–174. doi:10.1007/s00412-006-0048-6
37. Otto SP. The Evolutionary Consequences of Polyploidy. *Cell*. 2007;131: 452–462. doi:10.1016/j.cell.2007.10.022
38. Robert T, Vrielynck N, Mézard C, de Massy B, Grelon M. A new light on the meiotic DSB catalytic complex. *Semin Cell Dev Biol*. 2016;54: 165–176. doi:10.1016/j.semcdb.2016.02.025
39. Mézard C, Vignard J, Drouaud J, Mercier R. The road to crossovers: plants have their say. *Trends in Genetics*. 2007. pp. 91–99. doi:10.1016/j.tig.2006.12.007
40. Pawlowski WP. Chromosome organization and dynamics in plants. *Curr Opin Plant Biol*. 2010;13: 640–645. doi:10.1016/j.pbi.2010.09.015
41. Vrielynck N, Chambon A, Vezon D, Pereira L, Chelysheva L, De Muyt A, et al. A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science* (80- ). 2016;351: 939–943. doi:10.1126/science.aad5196
42. Hartung F, Wurz-Wildersinn R, Fuchs J, Schubert I, Suer S, Puchta H. The Catalytically Active Tyrosine Residues of Both SPO11-1 and SPO11-2 Are Required for Meiotic Double-Strand Break Induction in Arabidopsis. *Plant Cell*. 2007;19: 3090–3099. doi:10.1105/tpc.107.054817
43. Stacey NJ, Kuromori T, Azumi Y, Roberts G, Breuer C, Wada T, et al. Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J*. 2006;48: 206–216. doi:10.1111/j.1365-313X.2006.02867.x
44. Serrentino M-EE, Borde V. The spatial regulation of meiotic recombination hotspots:

- Are all DSB hotspots crossover hotspots? *Exp Cell Res.* Elsevier B.V.; 2012;318: 1347–1352. doi:10.1016/j.yexcr.2012.03.025
45. Lake CM, Hawley RS. Becoming a crossover-competent DSB. *Semin Cell Dev Biol.* 2016;54: 117–125. doi:10.1016/j.semcdb.2016.01.008
  46. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova G V. Genetic recombination is directed away from functional genomic elements in mice. *Nature.* 2012;485: 642–645. doi:10.1038/nature11089
  47. Liu Y, West SC. Happy Hollidays: 40th anniversary of the Holliday junction. *Nat Rev Mol Cell Biol.* 2004;5: 937–944. doi:10.1038/nrm1502
  48. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. *Nat Rev Genet.* Nature Publishing Group; 2013;14: 794–806. doi:10.1038/nrg3573
  49. Patel A, Horton JR, Wilson GG, Zhang X, Cheng X. Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev.* 2016;30: 257–265. doi:10.1101/gad.274928.115
  50. Ladias P, Markopoulos G, Lazaros L, Markoula S, Tzavaras T, Georgiou I. Holliday Junctions Are Associated with Transposable Element Sequences in the Human Genome. *J Mol Biol.* 2016;428: 658–667. doi:10.1016/j.jmb.2016.01.010
  51. Cromie GA, Hyppa RW, Taylor AF, Zakharyevich K, Hunter N, Smith GR. Single Holliday Junctions Are Intermediates of Meiotic Recombination. *Cell.* 2006;127: 1167–1178. doi:10.1016/j.cell.2006.09.050
  52. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair

- model for recombination. *Cell*. 1983;33: 25–35. Available: <http://www.ncbi.nlm.nih.gov/pubmed/6380756>
53. Pâques F, Haber JE. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*. 1999;63: 349–404. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10357855>
54. Sadava DE, Hillis DM, Heller HC, Berenbaum M. *Life: The Science of Biology* [Internet]. 7th ed. Sinauer Associates, Inc; 2014. Available: <http://www.sinauer.com/life-the-science-of-biology-769.html>
55. Phillips D, Jenkins G, Macaulay M, Nibau C, Wnetrzak J, Fallding D, et al. The effect of temperature on the male and female recombination landscape of barley. *New Phytol*. 2015;208: 421–429. doi:10.1111/nph.13548
56. Johnson DS. The History of the discovery of sexuality in plants. *Am Assoc Adv Sci*. 1914;39: 299–319.
57. Koszul R, Meselson M, Van Doninck K, Vandenhaute J, Zickler D. The Centenary of Janssens's Chiasmatype Theory. *Genetics*. 2012;191: 309–317. doi:10.1534/genetics.112.139733
58. Martini E, Diaz RL, Hunter N, Keeney S. Crossover Homeostasis in Yeast Meiosis. *Cell*. 2006;126: 285–295. doi:10.1016/j.cell.2006.05.044
59. Thacker D, Keeney S. PCH'ing Together an Understanding of Crossover Control. Copenhaver GP, editor. *PLoS Genet*. 2009;5: e1000576. doi:10.1371/journal.pgen.1000576
60. Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C. Genome-Wide

- Crossover Distribution in *Arabidopsis thaliana* Meiosis Reveals Sex-Specific Patterns along Chromosomes. Lichten M, editor. PLoS Genet. 2011;7: e1002354. doi:10.1371/journal.pgen.1002354
61. PARSONS PA. Evolutionary rates: effects of stress upon recombination. Biol J Linn Soc. 1988;35: 49–68. doi:10.1111/j.1095-8312.1988.tb00458.x
  62. Berchowitz LE, Copenhaver GP. Fluorescent *Arabidopsis* tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. Nat Protoc. 2008;3: 41–50. doi:10.1038/nprot.2007.491
  63. Choi K, Henderson IR. Meiotic recombination hotspots - A comparative view. Plant J. 2015;83: 52–61. doi:10.1111/tpj.12870
  64. Rhee SY, Somerville CR. Tetrad pollen formation in quartet mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall. Plant J. 1998;15: 79–88. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9744097>
  65. Dreissig S, Fuchs J, Cápál P, Kettles N, Byrne E, Houben A. Measuring Meiotic Crossovers via Multi-Locus Genotyping of Single Pollen Grains in Barley. Lichten M, editor. PLoS One. 2015;10: e0137677. doi:10.1371/journal.pone.0137677
  66. Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Genetics. 2002;162: 1401–1413.
  67. Wu J, Mizuno H, Hayashi-Tsugane M, Ito Y, Chiden Y, Fujisawa M, et al. Physical maps and recombination frequency of six rice chromosomes. Plant J. 2003;36: 720–730. doi:10.1046/j.1365-313X.2003.01903.x

68. and Analysis Consortium TCS. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005;437: 69–87. doi:10.1038/nature04072
69. Naranjo T, Valenzuela NT, Perera E. Chiasma Frequency Is Region Specific and Chromosome Conformation Dependent in a Rye Chromosome Added to Wheat. *Cytogenet Genome Res*. 2010;129: 133–142. doi:10.1159/000314029
70. Lukaszewski AJ, Kopecky D, Linc G. Inversions of chromosome arms 4AL and 2BS in wheat invert the patterns of chiasma distribution. *Chromosoma*. 2012;121: 201–208. doi:10.1007/s00412-011-0354-5
71. Mézard C. Meiotic recombination hotspots in plants. *Biochem Soc Trans*. 2006;34: 531–534. doi:10.1042/BST0340531
72. Webster MT, Hurst LD. Direct and indirect consequences of meiotic recombination: Implications for genome evolution. *Trends Genet*. Elsevier Ltd; 2012;28: 101–109. doi:10.1016/j.tig.2011.11.002
73. Dooner HK, Martínez-Férez IM. Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *Plant Cell*. 1997;9: 1633–1646. doi:10.1105/tpc.9.9.1633
74. Dooner HK. Genetic Fine Structure of the BRONZE Locus in Maize. *Genetics*. 1986;113: 1021–36. Available: /pmc/articles/PMC1202909/?report=abstract
75. He L, Dooner HK. Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proc Natl Acad Sci*. 2009;106: 8410–8416. doi:10.1073/pnas.0902972106

76. Saintenac C, Faure S, Remay A, Choulet F, Ravel C, Paux E, et al. Variation in crossover rates across a 3-Mb contig of bread wheat (*Triticum aestivum*) reveals the presence of a meiotic recombination hotspot. *Chromosoma*. 2011;120: 185–198. doi:10.1007/s00412-010-0302-9
77. Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, et al. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet*. Nature Publishing Group; 2013;45: 1327–1336. doi:10.1038/ng.2766
78. Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, et al. Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. Barsh GS, editor. *PLoS Genet*. 2012;8: e1002844. doi:10.1371/journal.pgen.1002844
79. Drouaud J, Khademian H, Giraut L, Zanni V, Bellalou S, Henderson IR, et al. Contrasted Patterns of Crossover and Non-crossover at Arabidopsis thaliana Meiotic Recombination Hotspots. *PLoS Genet*. 2013;9. doi:10.1371/journal.pgen.1003922
80. Yao H, Zhou Q, Li J, Smith H, Yandeu M, Nikolau BJ, et al. Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize. *Proc Natl Acad Sci*. 2002;99: 6157–6162. doi:10.1073/pnas.082562199
81. Yao H, Schnable PS. Cis-effects on Meiotic Recombination Across Distinct a1-sh2 Intervals in a Common Zea Genetic Background. *Genetics*. 2005;170: 1929–1944. doi:10.1534/genetics.104.034454
82. Smeds L, Mugal CF, Qvarnström A, Ellegren H. High-Resolution Mapping of Crossover and Non-crossover Recombination Events by Whole-Genome Re-sequencing of an Avian Pedigree. Henderson I, editor. *PLOS Genet*. 2016;12:

e1006044. doi:10.1371/journal.pgen.1006044

83. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet.* 2007;39: 1151–1155. doi:10.1038/ng2115
84. Henderson IR. Control of meiotic recombination frequency in plant genomes. *Curr Opin Plant Biol.* Elsevier Ltd; 2012;15: 556–561. doi:10.1016/j.pbi.2012.09.002
85. McVean G a T, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science.* 2004;304: 581–584. doi:10.1126/science.
86. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet.* 2008;40: 1124–1129. doi:10.1038/ng.213
87. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science.* 2005;310: 321–4. doi:10.1126/science.1117196
88. Liu S, Yeh C-T, Ji T, Ying K, Wu H, Tang HM, et al. Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome. Malik HS, editor. *PLoS Genet.* 2009;5: e1000733. doi:10.1371/journal.pgen.1000733
89. Pace JK, Feschotte C. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* 2007;17: 422–432. doi:10.1101/gr.5826307



90. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet.* 2007;41: 331–368. doi:10.1146/annurev.genet.40.110405.090448
91. Levin HL, Moran J V. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet.* 2011;12: 615–627. doi:10.1038/nrg3030
92. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8: 272–285. doi:10.1038/nrg2072
93. Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, et al. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* 2015;29: 1256–1270. doi:10.1101/gad.257840.114
94. Zhang L, Ma H. Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytol.* 2012;195: 248–263. doi:10.1111/j.1469-8137.2012.04143.x
95. Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in *Arabidopsis*. *Plant Cell.* 2015;27: 2427–2436. doi:10.1105/tpc.15.00391
96. Struhl K, Segal E. Determinants of nucleosome positioning. *Nat Struct Mol Biol.* 2013;20: 267–273. doi:10.1038/nsmb.2506
97. Baker K, Dhillon T, Colas I, Cook N, Milne I, Milne L, et al. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J.* 2015;84: 111–124. doi:10.1111/tpj.12963
98. Aliyeva-Schnorr L, Beier S, Karafiátová M, Schmutzer T, Scholz U, Doležel J, et al. Cytogenetic mapping with centromeric bacterial artificial chromosomes contigs shows

- that this recombination-poor region comprises more than half of barley chromosome 3H. *Plant J.* 2015;84: 385–394. doi:10.1111/tpj.13006
99. Winckler W. Comparison of Fine-Scale Recombination Rates in Humans and Chimpanzees. *Science* (80- ). 2005;308: 107–111. doi:10.1126/science.1105322
  100. Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. The Red Queen Model of Recombination Hotspots Evolution in the Light of Archaic and Modern Human Genomes. Coop G, editor. *PLoS Genet.* 2014;10: e1004790. doi:10.1371/journal.pgen.1004790
  101. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol.* 2002;5: 94–100. doi:10.1016/S1369-5266(02)00240-6
  102. Paape T, Zhou P, Branca A, Briskine R, Young N, Tiffin P. Fine-scale population recombination rates, hotspots, and correlates of recombination in the *Medicago truncatula* genome. *Genome Biol Evol.* 2012;4: 726–737. doi:10.1093/gbe/evs046
  103. Flint-Garcia S a, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol.* 2003;54: 357–374. doi:10.1146/annurev.arplant.54.031902.134907
  104. Rafalski A, Morgante M. Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet.* 2004;20: 103–111. doi:10.1016/j.tig.2003.12.002
  105. Awadalla P. The evolutionary genomics of pathogen recombination. *Nat Rev Genet.* 2003;4: 50–60. doi:10.1038/nrg964
  106. de Massy B. Initiation of Meiotic Recombination: How and Where? Conservation and

- Specificities Among Eukaryotes. *Annu Rev Genet.* 2013;47: 563–599.  
doi:10.1146/annurev-genet-110711-155423
107. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* (80- ). 2014;345: 1250092–1250092. doi:10.1126/science.1250092
108. Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet.* 2002;3: 429–441. Available: <http://dx.doi.org/10.1038/nrg817>
109. Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P. Detailed Recombination Studies Along Chromosome 3B Provide New Insights on Crossover Distribution in Wheat (*Triticum aestivum* L.). *Genetics.* 2008;181: 393–403. doi:10.1534/genetics.108.097469
110. Lukaszewski AJ, Curtis CA. Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor Appl Genet.* 1993;86. doi:10.1007/BF00223816
111. Shah MM, Hassan A. Distribution of genes and recombination on wheat homoeologous group 6 chromosomes: A synthesis of available information. *Mol Breed.* 2005;15: 43–53. doi:10.1007/s11032-004-1844-4
112. Paris JD, Haen KM, Gill BS. Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics.* 2000;154: 823–835.
113. Gill KS, Gill BS, Endo TR, Boyko E V. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics.* 1996;143: 1001–12. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8725245>

114. Gill KS, Gill BS, Endo TR, Taylor T. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics*. 1996;144: 1883–91. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8978071>
115. Naranjo T. Forcing the shift of the crossover site to proximal regions in wheat chromosomes. *Theor Appl Genet*. 2015;128: 1855–1863. doi:10.1007/s00122-015-2552-7
116. Wang G, Hyne V, Chao S, Gale MD, Henry Y, De Buyser J, et al. A comparison of male and female recombination frequency in wheat using RFLP maps of homoeologous group 6 and 7 chromosomes. *Theor Appl Genet*. 1995;91: 744–746. doi:10.1007/BF00220953
117. Koul KK, Nagpal R, Sharma A. Chromosome behaviour in the male and female sex mother cells of wheat (*Triticum aestivum* L.), oat (*Avena sativa* L.) and pearl millet (*Pennisetum americanum* (L.) Leeke). *Caryologia*. 2000;53: 175–183. doi:10.1080/00087114.2000.10589194
118. Guzy-Wróbelska J, Labocha-Pawlowska A, Kwasniewski M, Szarejko I. Different recombination frequencies in wheat doubled haploid populations obtained through maize pollination and anther culture. *Euphytica*. 2007;156: 173–183. doi:10.1007/s10681-007-9364-z
119. Fu TK, Sears ER. The Relationship between Chiasmata and Crossing over in TRITICUM AESTIVUM. *Genetics*. 1973;75: 231–46. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17248642>
120. QUEIROZ A, MELLO-SAMPAYO T, VIEGAS WS. Identification of low temperature stabilizing genes, controlling chromosome synapsis or recombination, in short arms of

- chromosomes from the homoeologous group 5 of *Triticum aestivum*. *Hereditas*. 2008;115: 37–41. doi:10.1111/j.1601-5223.1991.tb00344.x
121. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. *Science* (80- ). 2014;345: 1249721–1249721. doi:10.1126/science.1249721
122. Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, et al. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biol*. 2015;16: 188. doi:10.1186/s13059-015-0754-6
123. Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, et al. Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol*. 2014;15: 546. doi:10.1186/s13059-014-0546-4
124. Pingault L, Choulet F, Alberti A, Glover N, Wincker P, Feuillet C, et al. Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome Biol*. 2015;16: 29. doi:10.1186/s13059-015-0601-9
125. Rey M-D, Calderín MC, Prieto P. The use of the *ph1b* mutant to induce recombination between the chromosomes of wheat and barley. *Front Plant Sci*. 2015;6. doi:10.3389/fpls.2015.00160
126. RILEY R, CHAPMAN V. Genetic Control of the Cytologically Diploid Behaviour of Hexaploid Wheat. *Nature*. 1958;182: 713–715. doi:10.1038/182713a0
127. RILEY R, CHAPMAN V, KIMBER G. Genetic control of chromosome pairing in intergeneric hybrids with wheat. *Nature*. 1959;183: 1244–6. Available: <http://www.ncbi.nlm.nih.gov/pubmed/13657066>

128. CHAPMAN V, RILEY R. Homoeologous Meiotic Chromosome Pairing in *Triticum aestivum* in which Chromosome 5B is replaced by an Alien Homoeologue. *Nature*. 1970;226: 376–377. doi:10.1038/226376a0
129. Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, et al. Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature*. 2006;439: 749–752. doi:10.1038/nature04434
130. Martín AC, Shaw P, Phillips D, Reader S, Moore G. Licensing MLH1 sites for crossover during meiosis. *Nat Commun*. 2014;5: 4580. doi:10.1038/ncomms5580
131. Sears E. Cytogenetic studies with polyploid species of wheat. II. Additional chromosomal aberrations in *Triticum vulgare*. *Genetics*. 1944; 232–246.
132. MELLO-SAMPAYO T. Genetic Regulation of Meiotic Chromosome Pairing by Chromosome 3D of *Triticum aestivum*. *Nat New Biol*. 1971;230: 22–23. doi:10.1038/newbio230022a0
133. Mello-Sampayo T, Lorente R. The role of chromosome 3D in the regulation of meiotic pairing in hexaploid wheat. *EWAC Newsl*. 1968; 16–24.
134. Mello-Sampayo T, Canas P. Suppressors of meiotic chromosome pairing in common wheat. *Fourth Int Wheat Genet Symp*, Columbia, Missouri. 1973; 709–713.
135. Sears ER. A WHEAT MUTATION CONDITIONING AN INTERMEDIATE LEVEL OF HOMOEOLOGOUS CHROMOSOME PAIRING. *Can J Genet Cytol*. 1982;24: 715–719. doi:10.1139/g82-076
136. Sears ER. AN INDUCED MUTANT WITH HOMOEOLOGOUS PAIRING IN COMMON WHEAT. *Can J Genet Cytol*. 1977;19: 585–593. doi:10.1139/g77-063

137. Martinez-Perez E, Shaw P, Moore G. The Ph1 locus is needed to ensure specific somatic and meiotic centromere association. *Nature*. 2001;411: 204–207. doi:10.1038/35075597
138. Driscoll CJ. GENETIC SUPPRESSION OF HOMOEEOLOGOUS CHROMOSOME PAIRING IN HEXAPLOID WHEAT. *Can J Genet Cytol*. 1972;14: 39–42. doi:10.1139/g72-004
139. Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable recombination hotspots in birds. *Science (80- )*. 2015;350: 928–932. doi:10.1126/science.aad0843
140. Lam I, Keeney S. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science (80- )*. 2015;350: 932–937. doi:10.1126/science.aad0814
141. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*. 2009;457: 551–556. doi:10.1038/nature07723
142. Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F. Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). *Theor Appl Genet*. 2009;119: 1523–1537. doi:10.1007/s00122-009-1153-8
143. Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science (80- )*. 2014;345: 1251788–1251788. doi:10.1126/science.1251788
144. Šafař J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, et al. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from

- bread wheat. *Plant J.* 2004;39: 960–968. doi:10.1111/j.1365-313X.2004.02179.x
145. Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, et al. A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B. *Science* (80- ). 2008;322: 101–104. doi:10.1126/science.1161847
  146. Yu J-K, La Rota M, Kantety RV, Sorrells ME. EST derived SSR markers for comparative mapping in wheat and rice. *Mol Genet Genomics.* 2004;271. doi:10.1007/s00438-004-1027-3
  147. Prieto P, Shaw P, Moore G. Homologue recognition during meiosis is associated with a change in chromatin conformation. *Nat Cell Biol.* 2004;6: 906–908. doi:10.1038/ncb1168
  148. Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, et al. Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. *Proc Natl Acad Sci.* 2008;105: 6075–6080. doi:10.1073/pnas.0801521105
  149. Aragón-Alcaide L, Reader S, Beven A, Shaw P, Miller T, Moore G. Association of homologous chromosomes during floral development. *Curr Biol.* 1997;7: 905–8. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9382806>
  150. Jenczewski E, Mercier R, Macaisne N, Mézard C. Meiosis: Recombination and the Control of Cell Division. *Plant Genome Diversity Volume 2.* Vienna: Springer Vienna; 2013. pp. 121–136. doi:10.1007/978-3-7091-1160-4\_8
  151. Santos JL. The relationship between synapsis and recombination: two different views. *Heredity (Edinb).* 1999;82: 1–6. doi:10.1038/sj.hdy.6884870
  152. Higgins JD, Perry RM, Barakate A, Ramsay L, Waugh R, Halpin C, et al.



Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley. *Plant Cell*. 2012;24: 4096–4109. doi:10.1105/tpc.112.102483

153. Stack SM, Anderson LK. Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize. *Chromosome Res.* 2002;10: 329–45. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12199146>
154. Feldman M, Lupton F, Miller T. Wheats. Smartt J, Simmonds NW *Evol Crop*. 1995; 184–192.
155. Blake NK, Lehfeldt BR, Lavin M, Talbert LE. Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. *Genome / Natl Res Counc Canada = Génome / Cons Natl Rech Canada*. 1999;42: 351–60. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10231966>
156. Huang S, Sirikhachornkit A, Faris JD, Su X, Gill BS, Haselkorn R, et al. Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Mol Biol.* 2002;48: 805–20. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11999851>
157. Sears E. The Aneuploids of common Wheat. *Univ Missouri Agric Exp Stn*. 1954; 1–58.
158. Endo TR, Gill BS. The Deletion Stocks of Common Wheat. *J Hered* . 1996;87: 295–307. Available: <http://jhered.oxfordjournals.org/content/87/4/295.abstract>
159. Sears E, Sears L. The telocentric chromosomes of common wheat. *Proc 5th Wheat Genet Symp New Delhi*. 1978; 389–407.

160. Sharp PJ, Chao S, Desai S, Gale MD. The isolation, characterization and application in the Triticeae of a set of wheat RFLP probes identifying each homoeologous chromosome arm. *Theor Appl Genet.* 1989;78: 342–8. doi:10.1007/BF00265294
161. Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME. Optimizing parental selection for genetic linkage maps. *Genome.* 1993;36: 181–186. doi:10.1139/g93-024
162. Allouis S, Moore G, Chalhoub A, Bellec R, Sharp P, Faivre Rampant K, et al. Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm “Chinese Spring”. *Cereal Res Commun. Akadémiai Kiadó;* 2003;31: 331–338. Available: <http://www.jstor.org/stable/23786975>
163. Sears ER. Genetic Control of Chromosome Pairing in Wheat. *Annu Rev Genet.* 1976;10: 31–51. doi:10.1146/annurev.ge.10.120176.000335
164. Dover GA, Riley R. Inferences from Genetical Evidence on the Course of Meiotic Chromosome Pairing in Plants. *Philos Trans R Soc Lond B Biol Sci. The Royal Society;* 1977;277: 313–326. Available: <http://www.jstor.org/stable/2417717>
165. Dvorak J, Deal KR, Luo M-C. Discovery and Mapping of Wheat Ph1 Suppressors. *Genetics.* 2006;174: 17–27. doi:10.1534/genetics.106.058115
166. Miller TE, Chapman V. Aneuhaploids in bread wheat. *Genet Res.* 1976;28: 37. doi:10.1017/S0016672300016700
167. Viegas WS. The effect of B-chromosomes of rye on the chromosome association in F1 hybrids *Triticum aestivum* x *Secale cereale* in the absence of chromosomes 5B or 5D. *Theor Appl Genet.* 1980;56: 193–198. doi:10.1007/BF00295448

168. Denis J-B, BERNARD M, ARNOUX J, CAUDERON Y. Analyse statistique des configurations méiotiques lors de la création d'une série monosomique du blé tendre « Courtot ». *Agronomie*. 1982;2: 701–708. doi:10.1051/agro:19820803
169. Crismani W, Baumann U, Sutton T, Shirley N, Webster T, Spangenberg G, et al. Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. *BMC Genomics*. 2006;7: 267. doi:10.1186/1471-2164-7-267
170. Rustenholz C, Choulet F, Laugier C, Safár J, Simková H, Dolezel J, et al. A 3,000-Loci Transcription Map of Chromosome 3B Unravels the Structural and Functional Features of Gene Islands in Hexaploid Wheat. *PLANT Physiol*. 2011;157: 1596–1608. doi:10.1104/pp.111.183921
171. Sears ER. Nullisomic-Tetrasomic Combinations in Hexaploid Wheat. *Chromosome Manipulations and Plant Genetics*. Boston, MA: Springer US; 1966. pp. 29–45. doi:10.1007/978-1-4899-6561-5\_4
172. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28: 511–515. doi:10.1038/nbt.1621
173. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5: 621–628. doi:10.1038/nmeth.1226
174. Slater G, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6: 31. doi:10.1186/1471-2105-6-31
175. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped

- BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25: 3389–402. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9254694>
176. Boden S a., Langridge P, Spangenberg G, Able JA. TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1. *Plant J.* 2009;57: 487–497. doi:10.1111/j.1365-313X.2008.03701.x
  177. Aragon-Alcaide L, Beven A, Moore G, Shaw P. The use of vibratome sections of cereal spikelets to study anther development and meiosis. *Plant J.* 1998;14: 503–508. doi:10.1046/j.1365-313X.1998.00148.x
  178. Gerlach WL, Bedbrook JR. Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* 1979;7: 1869–1885. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC342353/>
  179. Das RM, Storey KG. Apical Abscission Alters Cell Polarity and Dismantles the Primary Cilium During Neurogenesis. *Science* (80- ). 2014;343: 200–204. doi:10.1126/science.1247521
  180. Devos K, Sorrells M, Anderson J, Miller T, Reader S, Lukaszewski A, et al. Chromosome aberrations in wheat nullisomic-tetrasomic and ditelosomic lines. *Cereal Res Commun.* 1999;27: 231–239.
  181. Barakate A, Higgins JD, Vivera S, Stephens J, Perry RM, Ramsay L, et al. The Synaptonemal Complex Protein ZYP1 Is Required for Imposition of Meiotic Crossovers in Barley. *Plant Cell.* 2014;26: 729–740. doi:10.1105/tpc.113.121269
  182. Al-Kaff N, Knight E, Bertin I, Foote T, Hart N, Griffiths S, et al. Detailed Dissection of the Chromosomal Region Containing the Ph1 Locus in Wheat *Triticum aestivum*: With

- Deletion Mutants and Expression Profiling. *Ann Bot.* 2008;101: 863–872.  
doi:10.1093/aob/mcm252
183. Boden SA, Shadiac N, Tucker EJ, Langridge P, Able JA. Expression and functional analysis of TaASY1 during meiosis of bread wheat (*Triticum aestivum*). *BMC Mol Biol.* 2007;8: 65. doi:10.1186/1471-2199-8-65
184. de Bustos A, Pérez R, Jouve N. Characterization of the gene Mre11 and evidence of silencing after polyploidization in *Triticum*. *Theor Appl Genet.* 2007;114: 985–999. doi:10.1007/s00122-006-0493-x
185. Sutton T, Whitford R, Baumann U, Dong C, Able JA, Langridge P. The Ph2 pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. *Plant J.* 2003;36: 443–56. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14617076>
186. Kempanna C, Riley R. Relationships between the genetic effects of deficiencies for chromosome III and V on meiotic pairing in *Triticum aestivum*. *Nature.* 1962; 1270–1273.
187. Sallee JP, Kimber G. An analysis of the pairing of wheat telocentric chromosomes. In: Ramanujam S, editor. *Proceedings of the 5th International Wheat Genetics Symposium*. New Delhi: Indian Society of Genetics and Plant Breeding; 1978. pp. 408–419.
188. Yousafzai FK, Al-Kaff N, Moore G. The molecular features of chromosome pairing at meiosis: The polyploid challenge using wheat as a reference. *Funct Integr Genomics.* 2010;10: 147–156. doi:10.1007/s10142-010-0171-6
189. Bassi FM, Kumar A, Zhang Q, Paux E, Huttner E, Kilian A, et al. Radiation hybrid

- QTL mapping of Tdes2 involved in the first meiotic division of wheat. *Theor Appl Genet.* 2013;126: 1977–1990. doi:10.1007/s00122-013-2111-z
190. Dvořák J. GENETIC VARIABILITY IN AEGILOPS SPELTOIDES AFFECTING HOMOEEOLOGOUS PAIRING IN WHEAT. *Can J Genet Cytol.* 1972;14: 371–380. doi:10.1139/g72-046
191. Beck M, Komis G, Muller J, Menzel D, Samaj J. Arabidopsis Homologs of Nucleus- and Phragmoplast-Localized Kinase 2 and 3 and Mitogen-Activated Protein Kinase 4 Are Essential for Microtubule Organization. *Plant Cell.* 2010;22: 755–771. doi:10.1105/tpc.109.071746
192. Zeng Q, Chen J-G, Ellis BE. AtMPK4 is required for male-specific meiotic cytokinesis in Arabidopsis. *Plant J.* 2011;67: 895–906. doi:10.1111/j.1365-313X.2011.04642.x
193. Franklin FCH, Higgins JD, Sanchez-Moran E, Armstrong SJ, Osman KE, Jackson N, et al. Control of meiotic recombination in Arabidopsis: role of the MutL and MutS homologues: Figure 1. *Biochem Soc Trans.* 2006;34: 542–544. doi:10.1042/BST0340542
194. Dion É, Li L, Jean M, Belzile F. An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. *Plant J.* 2007;51: 431–440. doi:10.1111/j.1365-313X.2007.03145.x
195. Sutani T, Yuasa T, Tomonaga T, Dohmae N, Takio K, Yanagida M. Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4. *Genes Dev.* 1999;13: 2271–83. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10485849>
196. Siddiqui NU. Mutations in Arabidopsis condensin genes disrupt embryogenesis,

- meristem organization and segregation of homologous chromosomes during meiosis. *Development*. 2003;130: 3283–3295. doi:10.1242/dev.00542
197. Shen Y, Tang D, Wang K, Wang M, Huang J, Luo W, et al. ZIP4 in homologous chromosome synapsis and crossover formation in rice meiosis. *J Cell Sci*. 2012;125: 2581–2591. doi:10.1242/jcs.090993
  198. Yuan L, Yang X, Ellis JL, Fisher NM, Makaroff CA. The Arabidopsis SYN3 cohesin protein is important for early meiotic events. *Plant J*. 2012;71: 147–160. doi:10.1111/j.1365-313X.2012.04979.x
  199. Hartung F, Suer S, Bergmann T, Puchta H. The role of AtMUS81 in DNA repair and its genetic interaction with the helicase AtRecQ4A. *Nucleic Acids Res*. 2006;34: 4438–4448. doi:10.1093/nar/gkl576
  200. Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D. A Plant Germline-Specific Integrator of Sperm Specification and Cell Cycle Progression. Copenhagen GP, editor. *PLoS Genet*. 2009;5: e1000430. doi:10.1371/journal.pgen.1000430
  201. Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D. The R2R3 MYB Transcription Factor DUO1 Activates a Male Germline-Specific Regulon Essential for Sperm Cell Differentiation in Arabidopsis. *Plant Cell*. 2011;23: 534–549. doi:10.1105/tpc.110.081059
  202. Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, Kammori M, et al. Regulation of Murine Telomere Length by Rtel. *Cell*. 2004;117: 873–886. doi:10.1016/j.cell.2004.05.026
  203. Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O’Neil NJ, Petalcorin MIR, et al. RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination.

Cell. 2008;135: 261–271. doi:10.1016/j.cell.2008.08.016

204. Hu Z, Cools T, Kalhorzadeh P, Heyman J, De Veylder L. Deficiency of the Arabidopsis Helicase RTEL1 Triggers a SOG1-Dependent Replication Checkpoint in Response to DNA Cross-Links. *Plant Cell Online*. 2015;27: 149–161. doi:10.1105/tpc.114.134312
205. Recker J, Knoll A, Puchta H. The Arabidopsis thaliana Homolog of the Helicase RTEL1 Plays Multiple Roles in Preserving Genome Stability. *Plant Cell Online*. 2014;26: 4889–4902. doi:10.1105/tpc.114.132472
206. Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Mol Plant*. 2015;8: 831–846. doi:10.1016/j.molp.2015.02.002
207. Ganai MW, Altmann T, Röder MS. SNP identification in crop plants. *Curr Opin Plant Biol*. 2009;12: 211–217. doi:10.1016/j.pbi.2008.12.009
208. Ganai MW, Polley A, Graner E-M, Plieske J, Wieseke R, Luerssen H, et al. Large SNP arrays for genotyping in crop plants. *J Biosci*. 2012;37: 821–828. doi:10.1007/s12038-012-9225-3
209. Paux E, Sourdille P, Mackay I, Feuillet C. Sequence-based marker development in wheat: Advances and applications to breeding. *Biotechnol Adv*. 2012;30: 1071–1088. doi:10.1016/j.biotechadv.2011.09.015
210. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nat Genet*. 2011;43: 956–963. doi:10.1038/ng.911



211. Kim N, Jeong Y-M, Jeong S, Kim G-B, Baek S, Kwon Y-E, et al. Identification of candidate domestication regions in the radish genome based on high-depth resequencing analysis of 17 genotypes. *Theor Appl Genet.* 2016;129: 1797–1814. doi:10.1007/s00122-016-2741-z
212. Kwong Q Bin, Teh CK, Ong AL, Heng HY, Lee HL, Mohamed M, et al. Development and Validation of a High-Density SNP Genotyping Array for African Oil Palm. *Mol Plant.* 2016;9: 1132–1141. doi:10.1016/j.molp.2016.04.010
213. Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet.* 2010;42: 1053–1059. doi:10.1038/ng.715
214. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol.* 2011;30: 105–111. doi:10.1038/nbt.2050
215. Winfield MO, Allen AM, BurrIDGE AJ, Barker GLA, Benbow HR, Wilkinson PA, et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J.* 2016;14: 1195–1206. doi:10.1111/pbi.12485
216. Jordan KW, Wang S, Lun Y, Gardiner L-J, MacLachlan R, Hucl P, et al. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 2015;16: 48. doi:10.1186/s13059-015-0606-4
217. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J.* 2014;12: 787–796. doi:10.1111/pbi.12183
218. Kobayashi F, Tanaka T, Kanamori H, Wu J, Katayose Y, Handa H. Characterization of

- a mini core collection of Japanese wheat varieties using single-nucleotide polymorphisms generated by genotyping-by-sequencing. *Breed Sci.* 2016;66: 213–225. doi:10.1270/jsbbs.66.213
219. Lai K, Lorenc MT, Lee HC, Berkman PJ, Bayer PE, Visendi P, et al. Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Plant Biotechnol J.* 2015;13: 97–104. doi:10.1111/pbi.12240
220. Edwards KJ, Reid AL, Coghill JA, Berry ST, Barker GLA. Multiplex single nucleotide polymorphism (SNP)-based genotyping in allohexaploid wheat using padlock probes. *Plant Biotechnol J.* 2009;7: 375–390. doi:10.1111/j.1467-7652.2009.00413.x
221. Bérard A, Le Paslier MC, Dardevet M, Exbrayat-Vinson F, Bonnin I, Cenci A, et al. High-throughput single nucleotide polymorphism genotyping in wheat (*Triticum* spp.). *Plant Biotechnol J.* 2009;7: 364–374. doi:10.1111/j.1467-7652.2009.00404.x
222. Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant J-P, et al. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol J.* 2010;8: 196–210. doi:10.1111/j.1467-7652.2009.00477.x
223. Akhunov E, Nicolet C, Dvorak J. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet.* 2009;119: 507–517. doi:10.1007/s00122-009-1059-5
224. Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, et al. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol J.* 2011;9: 1086–1099.

doi:10.1111/j.1467-7652.2011.00628.x

225. Cubizolles N, Rey E, Choulet F, Rimbert H, Laugier C, Balfourier F, et al. Exploiting the Repetitive Fraction of the Wheat Genome for High-Throughput Single-Nucleotide Polymorphism Discovery and Genotyping. *Plant Genome*. 2016;9. doi:10.3835/plantgenome2015.09.0078
226. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci*. 2013;110: 8057–8062. doi:10.1073/pnas.1217133110
227. Akhunov ED, Akhunova AR, Anderson OD, Anderson J a, Blake N, Clegg MT, et al. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics*. BioMed Central Ltd; 2010;11: 702. doi:10.1186/1471-2164-11-702
228. Clark RM. Estimating a Nucleotide Substitution Rate for Maize from Polymorphism at a Major Domestication Locus. *Mol Biol Evol*. 2005;22: 2304–2312. doi:10.1093/molbev/msi228
229. Rabinowicz PD. Differential methylation of genes and repeats in land plants. *Genome Res*. 2005;15: 1431–1440. doi:10.1101/gr.4100405
230. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 2012;6: 80–92. doi:10.4161/fly.19695
231. Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974;185: 862–4. Available: <http://www.ncbi.nlm.nih.gov/pubmed/4843792>

232. Poland JA, Brown PJ, Sorrells ME, Jannink J-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. Yin T, editor. *PLoS One*. 2012;7: e32253. doi:10.1371/journal.pone.0032253
233. Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet*. 2007;114: 1265–1275. doi:10.1007/s00122-007-0517-1
234. Bassil N V, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, et al. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria × ananassa*. *BMC Genomics*. 2015;16: 155. doi:10.1186/s12864-015-1310-1
235. Didion JP, Yang H, Sheppard K, Fu C-P, McMillan L, de Villena F, et al. Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. *BMC Genomics*. 2012;13: 34. doi:10.1186/1471-2164-13-34
236. Sorrells ME, Gustafson JP, Somers D, Chao S, Benscher D, Guedira-Brown G, et al. Reconstruction of the Synthetic W7984 × Opata M85 wheat reference population. *Genome*. 2011;54: 875–882. doi:10.1139/g11-054
237. Allen AM, Barker GLA, Wilkinson P, BurrIDGE A, Winfield M, Coghill J, et al. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol J*. 2013;11: 279–295. doi:10.1111/pbi.12009
238. Wang J, Luo M-C, Chen Z, You FM, Wei Y, Zheng Y, et al. *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic

- diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 2013;198: 925–937. doi:10.1111/nph.12164
239. Salomón-Torres R, González-Vizcarra VM, Medina-Basulto GE, Montaña-Gómez MF, Mahadevan P, Yaurima-Basaldúa VH, et al. Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. *Genet Mol Res.* 2015;14: 11848–11859. doi:10.4238/2015.October.2.18
240. Huang M-C, Chuang T-P, Chen C-H, Wu J-Y, Chen Y-T, Li L-H, et al. An integrated analysis tool for analyzing hybridization intensities and genotypes using new-generation population-optimized human arrays. *BMC Genomics.* 2016;17: 266. doi:10.1186/s12864-016-2478-8
241. Zhang X, Du R, Li S, Zhang F, Jin L, Wang H. Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinformatics.* 2014;15: 50. doi:10.1186/1471-2105-15-50
242. Fanciulli M, Petretto E, Aitman T. Gene copy number variation and common human disease. *Clin Genet.* 2010;77: 201–213. doi:10.1111/j.1399-0004.2009.01342.x
243. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet.* 2013;14: 125–138. doi:10.1038/nrg3373
244. Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. Copy Number Variation Affecting the Photoperiod-B1 and Vernalization-A1 Genes Is Associated with Altered Flowering Time in Wheat (*Triticum aestivum*). Hazen SP, editor. *PLoS One.* 2012;7: e33234. doi:10.1371/journal.pone.0033234

245. Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, et al. Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proc Natl Acad Sci.* 2013;110: 5241–5246. doi:10.1073/pnas.1220766110
246. Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. *Brief Funct Genomics.* 2014;13: 296–307. doi:10.1093/bfgp/elu016
247. Sutton T, Baumann U, Hayes J, Collins NC, Shi B-J, Schnurbusch T, et al. Boron-Toxicity Tolerance in Barley Arising from Efflux Transporter Amplification. *Science* (80- ). 2007;318: 1446–1449. doi:10.1126/science.1146853
248. Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, et al. A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. *New Phytol.* 2012;196: 282–291. doi:10.1111/j.1469-8137.2012.04243.x
249. Lin K, Smit S, Bonnema G, Sanchez-Perez G, de Ridder D. Making the difference: integrating structural variation detection tools. *Brief Bioinform.* 2015;16: 852–864. doi:10.1093/bib/bbu047
250. Ferreira A, Silva MF da, Silva L da C e, Cruz CD. Estimating the effects of population size and type on the accuracy of genetic maps. *Genet Mol Biol.* 2006;29: 187–192. doi:10.1590/S1415-47572006000100033
251. Erayman M. Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res.* 2004;32: 3546–3565. doi:10.1093/nar/gkh639
252. Gardner KA, Wittern LM, Mackay IJ. A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnol J.* 2016;14: 1406–1417. doi:10.1111/pbi.12504

253. Li H, Vikram P, Singh RP, Kilian A, Carling J, Song J, et al. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics*. 2015;16: 216. doi:10.1186/s12864-015-1424-5
254. Li C, Bai G, Chao S, Wang Z. A High-Density SNP and SSR Consensus Map Reveals Segregation Distortion Regions in Wheat. *Biomed Res Int*. 2015;2015: 1–10. doi:10.1155/2015/830618
255. Takumi S, Motomura Y, Iehisa JCM, Kobayashi F. Segregation distortion caused by weak hybrid necrosis in recombinant inbred lines of common wheat. *Genetica*. 2013;141: 463–470. doi:10.1007/s10709-013-9745-2
256. Kong L, Anderson JM, Ohm HW. Segregation distortion in common wheat of a segment of *Thinopyrum intermedium* chromosome 7E carrying Bdv3 and development of a Bdv3 marker. *Plant Breed*. 2009;128: 591–597. doi:10.1111/j.1439-0523.2009.01641.x
257. Shah R, Cavanagh CR, Huang BE. Computationally efficient map construction in the presence of segregation distortion. *Theor Appl Genet*. 2014;127: 2585–2597. doi:10.1007/s00122-014-2401-0
258. Dedryver F, Paillard S, Mallard S, Robert O, Trottet M, Nègre S, et al. Characterization of Genetic Components Involved in Durable Resistance to Stripe Rust in the Bread Wheat “Renan.” *Phytopathology*. 2009;99: 968–973. doi:10.1094/PHYTO-99-8-0968
259. Maia N. Obtention de blés tendres résistants au piétin-verse (*Cercospora herpotrichoides*) par croisements interspécifiques. 1967.
260. Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. Available: <http://www.repeatmasker.org>

261. Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, et al. Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces. *Plant Cell*. 2010;22: 1686–1701. doi:10.1105/tpc.110.074187
262. Wicker T, Matthews DE, Keller B. TREP: a database for Triticeae repetitive elements. *Trends Plant Sci*. 2002;7: 561–562. doi:10.1016/S1360-1385(02)02372-5
263. Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, et al. Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J*. 2006;48: 463–474. doi:10.1111/j.1365-313X.2006.02891.x
264. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
265. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033
266. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330
267. Perrier X, Flori A, Bonnot F. Methods for data analysis. *Genet Divers Cultiv Trop plants Sci Publ Inc CIRAD, Montpellier*. 2003; 31–63.
268. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006; Available: <http://darwin.cirad.fr/>
269. Wu Y, Bhat PR, Close TJ, Lonardi S. Efficient and Accurate Construction of Genetic



- Linkage Maps from the Minimum Spanning Tree of a Graph. Kruglyak L, editor. *PLoS Genet.* 2008;4: e1000212. doi:10.1371/journal.pgen.1000212
270. Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, et al. An integrative genetic linkage map of winter wheat ( *Triticum aestivum* L.). *TAG Theor Appl Genet.* 2003;107: 1235–1242. doi:10.1007/s00122-003-1361-6
271. Groos C, Robert N, Bervas E, Charmet G. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor Appl Genet.* 2003;106: 1032–1040. doi:10.1007/s00122-002-1111-1
272. Cadalen T, Boeuf C, Bernard S, Bernard M. An intervarietal molecular marker map in *Triticum aestivum* L. Em. Thell. and comparison with a map from a wide cross. *Theor Appl Genet.* 1997;94: 367–377. doi:10.1007/s001220050425
273. Thomas M, Pingault L, Poulet A, Duarte J, Throude M, Faure S, et al. Evolutionary history of Methyltransferase 1 genes in hexaploid wheat. *BMC Genomics.* 2014;15: 922. doi:10.1186/1471-2164-15-922
274. Iehisa JCM, Ohno R, Kimura T, Enoki H, Nishimura S, Okamoto Y, et al. A High-Density Genetic Map with Array-Based Markers Facilitates Structural and Quantitative Trait Locus Analyses of the Common Wheat Genome. *DNA Res.* 2014; 1–13. doi:10.1093/dnares/dsu020
275. Semagn K, Bjørnstad A, Skinnes H, Marøy AG, Tarkegne Y, William M. Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. *Genome.* 2006;49: 545–555. doi:10.1139/G06-002
276. Bing-Hua L, Jing-Yang D. A Dominant Gene for Male Sterility in Wheat. *Plant Breed.* 1986;97: 204–209. doi:10.1111/j.1439-0523.1986.tb01054.x

277. Gustafson P, Cao W, Somers DJ, Fedak G. A molecular marker closely linked to the region of Rht-D1c and Ms2 genes in common wheat ( *Triticum aestivum* ). *Genome*. 2009;52: 95–99. doi:10.1139/G08-097
278. Błaszczyk L, Goyeau H, Huang X-Q, Röder M, Stepień L, Chelkowski J. Identifying leaf rust resistance genes and mapping gene Lr37 on the microsatellite map of wheat. *Cell Mol Biol Lett*. 2004;9: 869–78. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15647803>
279. Li L, Jean M, Belzile F. The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. *Plant J*. 2006;45: 908–916. doi:10.1111/j.1365-313X.2006.02657.x
280. Opperman R, Emmanuel E, Levy A a. The effect of sequence divergence on recombination between direct repeats in *Arabidopsis*. *Genetics*. 2004;168: 2207–2215. doi:10.1534/genetics.104.032896
281. Li N, Stephens M. Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics*. 2003;165: 2213–2233. doi:10.1534/genetics.104.030692
282. Mayer KFX, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, et al. A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012. doi:10.1038/nature11543
283. Pan Q, Ali F, Yang X, Li J, Yan J. Exploring the Genetic Characteristics of Two Recombinant Inbred Line Populations via High-Density SNP Markers in Maize. *PLoS One*. 2012;7: 1–9. doi:10.1371/journal.pone.0052777
284. Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE,

- et al. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc Natl Acad Sci.* 2015;112: 201413864. doi:10.1073/pnas.1413864112
285. Duret L, Marais G, Biémont C. Transposons but Not Retrotransposons Are Located Preferentially in Regions of High Recombination Rate in *Caenorhabditis elegans*. *Genetics.* 2000;156: 1661–1669. Available: <http://genetics.org/content/156/4/1661.abstract>
286. Rizzon C, Marais G, Gouy M, Biemont C. Recombination Rate and the Distribution of Transposable Elements in the *Drosophila melanogaster* Genome. *Genome Res.* 2002;12: 400–407. doi:10.1101/gr.210802
287. Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. Population Genomics of Transposable Elements in *Drosophila*. *Annu Rev Genet.* 2014;48: 561–581. doi:10.1146/annurev-genet-120213-092359
288. Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. *Genes Dev.* 2015;29: 2183–2202. doi:10.1101/gad.270876.115
289. Smit AFA. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 1993;21: 1863–1872. doi:10.1093/nar/21.8.1863
290. Stumpf MPH, McVean GAT. Estimating recombination rates from population-genetic data. *Nat Rev Genet.* 2003;4: 959–968. doi:10.1038/nrg1227
291. Darrier B, Rimbert H, Navarro J, Cubizolles N, Choulet F, Leveugle M, et al. High throughput SNP discovery and genotyping in hexaploid wheat.

292. Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet.* 2004;36: 700–706. doi:10.1038/ng1376
293. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Vienna, Austria; 2008. Available: <http://www.r-project.org>
294. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.* 2015;12: 115–121. doi:10.1038/nmeth.3252
295. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2: 28–36. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7584402>
296. Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, et al. A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation. *Cell.* 2011;144: 719–731. doi:10.1016/j.cell.2011.02.009
297. Tischfield SE, Keeney S. Scale matters : the spatial correlation of yeast meiotic DNA breaks with histone H3 trimethylation is driven largely by independent. *Cell Cycle.* 2012;11: 1496–1503. doi:10.4161/cc.19733
298. Comeron JM, Ratnappan R, Bailin S. The Many Landscapes of Recombination in *Drosophila melanogaster*. *PLoS Genet.* 2012;8: 33–35. doi:10.1371/journal.pgen.1002905
299. Auton A, Rui Li Y, Kidd J, Oliveira K, Nadel J, Holloway JK, et al. Genetic

- Recombination Is Targeted towards Gene Promoter Regions in Dogs. Henderson I, editor. *PLoS Genet.* 2013;9: e1003984. doi:10.1371/journal.pgen.1003984
300. Zhao D, Ferguson AA, Jiang N. What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim Biophys Acta - Gene Regul Mech.* 2016;1859: 366–380. doi:10.1016/j.bbagr.2015.12.005
301. Wang J, Fan HC, Behr B, Quake SR. Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. *Cell.* 2012;150: 402–412. doi:10.1016/j.cell.2012.06.030
302. Lu S, Zong C, Fan W, Yang M, Li J, Chapman AR, et al. Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing. *Science (80- ).* 2012;338: 1627–1630. doi:10.1126/science.1229112
303. Deschamps S, Llaca V, May GD. Genotyping-by-Sequencing in Plants. *Biology (Basel).* 2012;1: 460–483. doi:10.3390/biology1030460
304. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature.* 2008;454: 479–485. doi:10.1038/nature07135
305. Yandea-Nelson MD, Zhou Q, Yao H, Xu X, Nikolau BJ, Schnable PS. MuDR Transposase Increases the Frequency of Meiotic Crossovers in the Vicinity of a Mu Insertion in the Maize *al* Gene. *Genetics.* 2005;169: 917–929. doi:10.1534/genetics.104.035089
306. Wijnker E, Velikkakam James G, Ding J, Becker F, Klasen JR, Rawat V, et al. The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife.* 2013;2. doi:10.7554/eLife.01426

307. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science* (80- ). 2010;327: 836–840. doi:10.1126/science.1183439
308. Feschotte C, Wessler SR. Mariner-like transposases are widespread and diverse in flowering plants. *Proc Natl Acad Sci*. 2002;99: 280–285. doi:10.1073/pnas.022626699
309. Pratto F, Brick K, Khil P, Smagulova F, Petukhova G V., Camerini-Otero RD. Recombination initiation maps of individual human genomes. *Science* (80- ). 2014;346: 1256442–1256442. doi:10.1126/science.1256442
310. Muñoz-Fuentes V, Di Rienzo A, Vilà C. Prdm9, a Major Determinant of Meiotic Recombination Hotspots, Is Not Functional in Dogs and Their Wild Relatives, Wolves and Coyotes. Dawson DS, editor. *PLoS One*. 2011;6: e25498. doi:10.1371/journal.pone.0025498
311. Heil CSS, Noor MAF. Zinc Finger Binding Motifs Do Not Explain Recombination Rate Variation within or between Species of *Drosophila*. Singh N, editor. *PLoS One*. 2012;7: e45055. doi:10.1371/journal.pone.0045055
312. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8: 2281–2308. doi:10.1038/nprot.2013.143
313. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science* (80- ). 2014;346: 1258096–1258096. doi:10.1126/science.1258096
314. Peciña A, Smith KN, Mézard C, Murakami H, Ohta K, Nicolas A. Targeted stimulation of meiotic recombination. *Cell*. 2002;111: 173–84. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12408862>

315. Yelina N, Diaz P, Lambing C, Henderson IR. Epigenetic control of meiotic recombination in plants. *Sci China Life Sci.* 2015; doi:10.1007/s11427-015-4811-x
316. Higgins JD, Osman K, Jones GH, Franklin FCH. Factors Underlying Restricted Crossover Localization in Barley Meiosis. *Annu Rev Genet.* 2014;48: 29–47. doi:10.1146/annurev-genet-120213-092509
317. R Core Development Team. R: a language and environment for statistical computing, 3.2.1. Document freely available on the internet at: <http://www.r-project.org>. Vienna, Austria: R Foundation for Statistical Computing; 2015. doi:10.1017/CBO9781107415324.004
318. Sosnowski O, Charcosset A, Joets J. BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics.* 2012;28: 2082–2083. doi:10.1093/bioinformatics/bts313

## Annexe

---

### A. Cartographie génome entier du blé tendre

Les différents travaux de génotypage ont été réalisés à l'aide de la puce AXIOM 420K (Affymetrix) développée dans le cadre du projet BreedWheat (E Paux, communication personnelle) qui possède 423 385 marqueurs SNP (Single Nucleotide Polymorphism) dont 115 147 sont issus directement du projet BreedWheat, 280 354 du consortium international IWGSC, 13 670 d'une précédente validation sur une puce Illumina [217], 4 815 de l'INRA, 5 155 de l'université de Bristol et 4 120 de Biogemma.

A l'aide des données de génotypage issues de deux algorithmes Affymetrix différents, j'ai pu réaliser la cartographie génome entier du blé tendre. Afin de réaliser cette cartographie, j'ai d'abord élaboré un script afin de traiter les matrices de génotypage de manière automatique, puis j'ai procédé à la réalisation des cartes génétiques avec MSTmap [269].

### B. Script de traitement des données de génotypage

Ce script rédigé sous R [317] a pour but de réaliser les cartographies génétiques avec les marqueurs les plus robustes, chromosome par chromosome. L'assignation des marqueurs génétiques aux différents chromosomes a été réalisée en amont lors de leur détection. Ce script permet, en plusieurs étapes de sélectionner les marqueurs chromosome par chromosome, d'éliminer les marqueurs hétérozygotes, présentant trop (>15%) de valeurs manquantes ou présentant une distorsion de ségrégation (test statistique de  $\chi^2 < 1\%$ ) ainsi que les individus présentant plus de 20% de valeurs manquantes. A la suite de l'analyse plusieurs fichiers de sortie sont générés afin de permettre le suivi.



### **C. Cartographique génétique.**

Afin de réaliser la cartographie génétique et de placer le plus possible de marqueurs, plusieurs analyses ont dû être effectuées. Tout d'abord, les marqueurs les plus robustes ont été cartographiés chromosome par chromosome. Ces marqueurs, assignés aux différents chromosomes grâce aux lignées aneuploïdes, sont sélectionnés en fonction de leur facilité d'analyse (nuages non ambigus pour les paramètres statistiques affymetrix ; parent homozygote), du faible nombre de données manquantes (<5%) et d'une absence de biais de ségrégation ( $P(\chi^2) > 0,01$ ). A la suite de cette première cartographie, un marqueur pour chaque bin génétique de chaque chromosome a été conservé afin d'obtenir un squelette à la fois dense, informatif et solide pour dans une deuxième étape cartographier les marqueurs restant c'est-à-dire des marqueurs soit robustes mais mal assignés aux chromosomes, soit hétérozygotes pour l'un des parents, soit qui possèdent des valeurs manquantes pour les parents et qui après traitement manuel ont pu être rajoutés à l'analyse. Enfin, dans une troisième étape, les marqueurs présentant un biais de ségrégation important ( $P(\chi^2) < 0,01$ ) ont été rajoutés.

### **D. Population Chinese Spring X Renan**

La population Chinese Spring X Renan est une population sélectionnée en Single Seed Descent en 6<sup>ème</sup> génération. Cette population n'a fait l'objet d'aucune sélection volontaire. Elle comporte 1270 individus. Le parent maternel de la population est la variété Chinese Spring qui est un blé tendre non cultivé utilisé comme de référence internationale pour le séquençage du génome. Chinese Spring est un blé de printemps de grande taille, sensible à de très nombreuses maladies. Le parent mâle de la population est la variété Renan sélectionnée par l'INRA et déposée au catalogue en 1989. Il s'agit d'un blé d'hiver de petite taille qui cumule des résistances aux rouilles, au piétin-verse, à la septoriose et à la fusariose. Elle possède une très haute valeur boulangère ainsi qu'une bonne tolérance au froid [14].

## **E. Population ITMI**

La population ITMI (International Triticeae Mapping Initiative) est une population issue du croisement entre les parents W7984, blé tendre synthétique dérivé de l'hybridation entre *Ae. tauschii* [DD] et la variété de blé dur Altar [AABB], et Opata M85 [AABBDD] blé cultivé Mexicain. Cette population a été créée afin d'obtenir un fort taux de polymorphisme pour permettre la construction aisée de cartes génétiques denses, en particulier au niveau du génome D habituellement peu polymorphe. Dans le cadre de l'étude il s'agit de la population ITMI SynOpDH pour synthétique X Opata Double Haploïde. Cette population est composée de 215 individus haploïdes doublés.

## **F. Projection des cartes et addition de marqueurs par bioinformatique**

L'analyse des puces de la population Chinese Spring X Renan, s'est faite à travers deux algorithmes différents ne fournissant pas la même qualité de données. Ainsi deux cartes Chinese Spring X Renan ont été calculées : la carte PolyHigh dont les données sont de très bonne qualité et la carte Flat dont la distance génétique des chromosomes avoisine les 1000cM et qui potentiellement possède des erreurs de génotypage.

La projection des cartographies des populations Chinese Spring X Renan PolyHigh, Flat et ITMI a été réalisée à l'aide de Biomercator V4.0 [318]. Cette projection a été établie en utilisant la cartographie Chinese Spring X Renan PolyHigh comme carte de référence.

L'addition de marqueurs par bioinformatique a été réalisée par attribution d'une distance génétique au scaffold qui porte au moins un marqueur placé génétiquement. Lorsque plusieurs marqueurs placés génétiquement sur un scaffold étaient présents, un choix aléatoire a été effectué. Finalement les marqueurs présents sur ces scaffolds et non cartographiés génétiquement se sont vu attribuer la position génétique du scaffold.





## Résumé

Mieux connaître les facteurs qui gouvernent l'apparition des événements de recombinaison (crossing-overs ; CO) chez le blé tendre (*Triticum aestivum* L.) est primordial car ce processus est l'outil principal du sélectionneur pour permettre le brassage de la diversité génétique et l'introgession de régions d'intérêt dans des variétés agronomiques. L'utilisation de techniques de cytogénétique développées sur l'orge a permis de comparer la mise en place de la synapse lors de la méiose chez des lignées de blé tendre délétées de tout ou partie du bras long du chromosome 3B et qui avaient été préalablement montrées comme présentant un nombre de chiasmas réduit par rapport à la variété euploïde. Les analyses cytogénétiques couplées à des études bioinformatiques de la séquence montrent que le synapsis a lieu quasiment normalement chez les mutants et que la délétion de certains gènes connus comme impactant le déroulement de la méiose pourrait expliquer le phénotype observé. De plus, le développement de ressources génomiques (SNP, séquence) à destination des sélectionneurs a permis la réalisation de cartes génétiques haute densité des 21 chromosomes ancrées sur la séquence du génome. Tous les chromosomes montrent le même profil de recombinaison avec un accroissement dans les parties distales et une réduction drastique dans les parties centromériques et péri-centromériques. L'exploitation de plus de 250 CO localisés dans des fenêtres de moins de 25 kb sur le chromosome 3B utilisé comme modèle pour l'étude de l'impact de la séquence sur la recombinaison, montre que les profils de recombinaison ancestrale et actuelle sont conservés et que les CO ont lieu préférentiellement dans les parties promotrices des gènes exprimés en méiose ce qui suggère que la conformation chromatinienne influence la recombinaison. Finalement, ces données ont aussi été l'opportunité de détecter des motifs liés à la recombinaison qui présentent des similarités avec celui ciblé par la protéine PRDM9 qui conduit à la recombinaison chez l'humain. Cela suggère que les mécanismes de contrôle de la recombinaison sont conservés chez les eucaryotes.

Mots clés : recombinaison ; méiose ; *Triticum aestivum* ; blé tendre ; crossing-over ; cartes génétiques ; cytogénétique.

## Abstract

Better knowledge of the factors that drive recombination (crossovers; COs) in bread wheat (*Triticum aestivum* L.) is of main interest since this process is the main tool allowing breeders to admix the genetic diversity and to introgress regions of interest in agronomic varieties. We used cytogenetic techniques previously developed on barley to compare the establishment of synapsis during meiosis in deletion lines missing part or whole of the long arm of chromosome 3B of bread wheat and which were previously shown as having a reduced chiasmata number compared to euploid varieties. Cytogenetic analysis combined with bioinformatics studies showed that the synapsis occurs almost normally in mutants and that deletion of some genes known to impact meiosis behavior may explain the observed phenotype. In addition, development of genomic resources (SNPs, sequence) for wheat breeders allowed simultaneous elaboration of high density genetic maps for the 21 chromosomes anchored on genome sequence. All chromosomes present the same recombination pattern with an increase of recombination in the distal parts and reduction in centromeric/péricentromeric regions of the chromosomes. Analysis of more than 250 COs mapped in windows lower than 25kb located on chromosome 3B used as model to study the impact of sequence features on recombination showed that current and ancestral recombination patterns are conserved and that COs preferentially occur in the promoter part of gene expressed in meiosis suggesting that chromatin conformation impacts recombination. Finally, these data were the opportunity to detect recombination-associated motif which presents similarities with the motif targeted by the PRDM9 protein driving recombination in human. This suggests that the control of recombination mechanisms is conserved among eukaryotes.

Key words: recombination; meiosis; *Triticum aestivum*; bread wheat; crossovers; genetic maps; cytogenetics.